PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Assistant Commissioner for Patents
United States Patent and Trademark Office
Box PCT
Washington, D.C.20231
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 18 October 2000 (18.10.00)	
International application No. PCT/GB00/01002	**Applicant's or agent's file reference** P6478WO ATM
International filing date (day/month/year) 17 March 2000 (17.03.00)	**Priority date** (day/month/year) 17 March 1999 (17.03.99)
Applicant UDEN, Mark et al	

1. The designated Office is hereby notified of its election made:

[X] in the demand filed with the International Preliminary Examining Authority on:

16 August 2000 (16.08.00)

[ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was

[ ] was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer

Zakaria EL KHODARY

Telephone No.: (41-22) 338.83.38

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/86, 9/00, 9/22, 7/04, 5/10 // A61P 31/18

A1

(11) International Publication Number: WO 00/55341

(43) International Publication Date: 21 September 2000 (21.09.00)

(21) International Application Number: PCT/GB00/01002

(22) International Filing Date: 17 March 2000 (17.03.00)

(30) Priority Data:
9906177.2 17 March 1999 (17.03.99) GB

(71) Applicant *(for all designated States except US)*: OXFORD BIOMEDICA (UK) LIMITED [GB/GB]; Medawar Centre, The Oxford Science Park, Robert Robinson Avenue, Oxford OX4 4GA (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: UDEN, Mark [GB/GB]; Flat 2, Finsbury Park, 17 Sommerfield Road, London N4 2JN (GB). MITROPHANOUS, Kyriacos [GR/US]; 85 Warwick Street, Oxford OX4 1SZ (US).

(74) Agents: MASCHIO, Antonio et al.; D Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: ANTI-VIRAL VECTORS

(57) Abstract

A viral vector production system is provided which system comprises: (i) a viral genome comprising at least one first nucleotide sequence encoding a gene product capable of binding to and effecting the cleavage, directly or indirectly, of a second nucleotide sequence, or transcription product thereof, encoding a viral polypeptide required for the assembly of viral particles; (ii) a third nucleotide sequence encoding said viral polypeptide required for the assembly of the viral genome into viral particles, which third nucleotide sequence has a different nucleotide sequence to the second nucleotide sequence such that said third nucleotide sequence, or transcription product thereof, is resistant to cleavage directed by said gene product; wherein at least one of the gene products is an external guide sequence capable of binding to and effecting the cleavage by RNase P of the second nucleotide sequence. The viral vector production system may be used to produce viral particles for use in treating or preventing viral infection.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
P006478WO ATM		
International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/GB00/01002	17/03/2000	17/03/1999
International Patent Classification (IPC) or national classification and IPC C12N15/86		
Applicant OXFORD BIOMEDICA (UK) LIMITED et al.		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 5 sheets, including this cover sheet.

   ☐ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☐ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   - VI ☐ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☐ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
16/08/2000	01.06.2001
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Wimmer, G Telephone No. +49 89 2399 7347

Form PCT/IPEA/409 (cover sheet) (January 1994)

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/GB00/01002

I. Basis of the report

1. With regard to the **elements** of the international application *(Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17))*:
Description, pages:

1-30 as originally filed

Claims, No.:

1-23 as originally filed

Drawings, sheets:

1/14-14/14 as originally filed

Sequence listing part of the description, pages:

1-40 (SEQ ID NOs. 1-73), as originally filed

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

These elements were available or furnished to this Authority in the following language: , which is:

- ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).
- ☐ the language of publication of the international application (under Rule 48.3(b)).
- ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

- ☒ contained in the international application in written form.
- ☐ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☒ furnished subsequently to this Authority in computer readable form.
- ☒ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☒ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:

- ☐ the description, pages:
- ☐ the claims, Nos.:
- ☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

   (Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	1-23
	No:	Claims	
Inventive step (IS)	Yes:	Claims	
	No:	Claims	1-23
Industrial applicability (IA)	Yes:	Claims	1-23
	No:	Claims	

2. Citations and explanations
   see separate sheet

Re Item V

Reasoned statement under Art. 35(2) PCT with regard to novelty, inventive step or industrial applicability.

The application does not meet the requirements of Art. 33 PCT since claims 1-23 do not appear to contain an inventive step.

1) Reference is made to the following documents (the document numbering corresponds to their order of citation in the international search report):

D1: WO 97 20060 A (UNIV JOHNS HOPKINS MED) 5 June 1997 (1997-06-05)

D2: YUAN Y ET AL: 'Targeted cleavage of mRNA by human RNase P' PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 89, no. 89, September 1992 (1992-09), pages 8006-8010, XP002104826 ISSN: 0027-8424 cited in the application

Novelty under Art. 33(2) PCT.

2) Although the prior art discloses vector systems which utilize ribozymes to directly cleave wild-type viral RNA, no documents describe such a system with External Guide Sequences to cause indirect cleavage of RNA. The systems, methods and viral particles of claims 1-23 are therefore novel.

Inventive Step under Art. 33(3) PCT.

3) Document D1 can be regarded as the closest prior art. In this document, vector systems are described in which ribozymes are used to target and cleave viral RNA. Specific examples of D1 include HIV vector systems, which include one or more ribozymes targeting sequences within the wild-type viral RNA. Furthermore, said ribozymes do not cleave within sequences of the described vectors, since the vector sequences, although still supporting packaging of the vector, have been altered from wild-type sequences so that the ribozyme recognition sites are absent.

Through this, the inventions of D1 and that of the present application are different in that D1 only describes the use of ribozymes, but not of External Guide Sequences to direct cleavage of viral RNA.

The technical problem at the basis of the present application was therefore to find means of targeting and cleaving viral RNA alternative to the use of ribozymes.

The solution, the use of specifically designed External Guide Sequences, however, cannot be viewed as being inventive.
Document D5 describes the successful construction of External Guide Sequences, for directed cleavage of a reporter gene RNA in human cells. Furthermore, D5 explains how to construct further EG sequences to cleave any desired RNA, and proposes this technique as a "general technique for gene inactivation". The document also states that this approach "has an advantage over techniques of gene inactivation that involve antisense RNA or other, exogenous ribozymes" (pg. 8010).
The person skilled in the art would therefore try to use the technique of D5, in the method of D1. He would therefore try, with reasonable chance of success, to construct sequences for EGS-directed cleavage of viral RNA, preferably for regions within the viral genome already targeted by the method of D1 (i.e., the HIV gag-pol sequence). The skilled person would use these sequences in place of, or additional to, the ribozyme coding sequences within the constructs of D1, and arrive at the invention of the present application.

Consequently, these methods and entities, and accordingly also claims 1-23, lack an inventive step.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

ANTI-VIRAL VECTORS

Field of the Invention

The present invention relates to novel viral vectors capable of delivering anti-viral inhibitory RNA molecules to target cells.

Background to the Invention

The application of gene therapy to the treatment of AIDS and HIV infection has been discussed widely (Lever, 1995). The types of therapeutic gene proposed usually fall into one of two broad categories. In the first the gene encodes protein products that inhibit the virus in a number of possible ways. One example of such a protein is the RevM10 derivative of the HIV Rev protein. The RevM10 protein acts as a transdominant negative mutant and so competitively inhibits Rev function in the virus. Like many of the protein-based strategies, the RevM10 protein is a derivative of a native HIV protein. While this provides the basis for the anti-HIV effect, it also has serious disadvantages. In particular, this type of strategy demands that in the absence of the virus there is little or no expression of the gene. Otherwise, healthy cells harbouring the gene become a target for the host cytotoxic T lymphocyte (CTL) system, which recognises the foreign protein. The second broad category of therapeutic gene circumvents these CTL problems. The therapeutic gene encodes inhibitory RNA molecules; RNA is not a target for CTL recognition.

There are several types of inhibitory RNA molecules known: anti-sense RNA, ribozymes, competitive decoys and external guide sequences (EGSs).

External guide sequences, first identified by Forster and Altman (1990), are RNA sequences that are capable of directing the cellular protein RNase P to cleave a particular RNA sequence. *In vivo*, they are found as part of precursor tRNAs where they function to direct cleavage by the cellular riboprotein RNase P *in vivo* of the tRNA precursor to form mature tRNA. However, in principle, any RNA can be targeted by a custom-designed EGS RNA for specific cleavage by RNase P *in vitro* or *in vivo*. For example, Yuan *et al.* (1992)

demonstrate a reduction in the levels of chloramphenicol activity in cells in tissue culture as a result of introducing an appropriately designed EGS.

In recent years a number of laboratories have developed retroviral vector systems based on HIV. In the context of anti-HIV gene therapy these vectors have a number of advantages over the more conventional murine based vectors such as murine leukaemia virus (MLV) vectors. Firstly, HIV vectors would target precisely those cells that are susceptible to HIV infection. Secondly, the HIV-based vector would transduce cells such as macrophages that are normally refractory to transduction by murine vectors. Thirdly, the anti-HIV vector genome would be propagated through the CD4+ cell population by any virus (HIV) that escaped the therapeutic strategy. This is because the vector genome has the packaging signal that will be recognised by the viral particle packaging system. These various attributes make HIV-vectors a powerful tool in the field of anti-HIV gene therapy.

A combination of inhibitory RNA molecules and an HIV-based vector would be attractive as a therapeutic strategy. However, until now this has not been possible. Vector particle production takes place in producer cells which express the packaging components of the particles and package the vector genome. The inhibitory RNA sequences that are designed to destroy the viral RNA would therefore also interrupt the expression of the components of the HIV-based vector system during vector production. The present invention aims to overcome this problem.

Summary of the Invention

It is therefore an object of the invention to provide a system and method for producing viral particles, in particular HIV particles, which carry nucleotide constructs encoding inhibitory RNA molecules such as external guide sequences, optionally together with other classes of inhibitory RNA molecules such as ribozymes and/or antisense RNAs directed against a corresponding virus, such as HIV, within a target cell, that overcomes the above-mentioned problems. The system includes both a viral genome encoding the inhibitory RNA molcules and nucleotide constructs encoding the components required for packaging the viral genome in a producer cell. However, in contrast to the prior art, although the packaging components have substantially the same amino acid sequence as the corresponding

components of the target virus, the inhibitory RNA molecules do not affect production of the viral particles in the producer cells because the nucleotide sequence of the packaging components used in the viral system have been modified to prevent the inhibitory RNA molecules from effecting cleavage or degradation of the RNA transcripts produced from the constructs. Such a viral particle may be used to treat viral infections, in particular HIV infections.

Accordingly the present invention provides a viral vector system comprising:

(i) a first nucleotide sequence encoding an external guide sequence capable of binding to and effecting the cleavage by RNase P of a second nucleotide sequence, or transcription product thereof, encoding a viral polypeptide required for the assembly of viral particles; and

(ii) a third nucleotide sequence encoding said viral polypeptide required for the assembly of viral particles, which third nucleotide sequence has a different nucleotide sequence to the second nucleotide sequence such that the third nucleotide sequence, or transcription product thereof, is resistant to cleavage directed by the external guide sequence.

Preferably, said system further comprises at least one further first nucleotide sequence encoding a gene product capable of binding to and effecting the cleavage, directly or indirectly, of a second nucleotide sequence, or transcription product thereof, encoding a viral polypeptide required for the assembly of viral particles, wherein the gene product is selected from an external guide sequence, a ribozyme and an anti-sense ribonucleic acid.

In another aspect, the present invention provides a viral vector production system comprising:

(i) a viral genome comprising at least one first nucleotide sequence encoding a gene product capable of binding to and effecting the cleavage, directly or indirectly, of a second nucleotide sequence, or transcription product thereof, encoding a viral polypeptide required for the assembly of viral particles;

(ii) a third nucleotide sequence encoding said viral polypeptide required for the assembly of the viral genome into viral particles, which third nucleotide sequence has a different nucleotide sequence to the second nucleotide sequence such that said third

nucleotide sequence, or transcription product thereof, is resistant to cleavage directed by said gene product;

wherein at least one of the gene products is an external guide sequence capable of binding to and effecting the cleavage by RNase P of the second nucleotide sequence.

Preferably, in addition to an external guide sequence, at least one gene product is selected from a ribozyme and an anti-sense ribonucleic acid, preferably a ribozyme.

Preferably, the viral vector is a retroviral vector, more preferably a lentiviral vector, such as an HIV vector. The second nucleotide sequence and the third nucleotide sequences are typically from the same viral species, more preferably from the same viral strain. Generally, the viral genome is also from the same viral species, more preferably from the same viral strain.

In the case of retroviral vectors, the polypeptide required for the assembly of viral particles is selected from gag, pol and env proteins. Preferably at least the gag and pol sequences are lentiviral sequences, more preferably HIV sequences. Alternatively, or in addition, the env sequence is a lentiviral sequence, more preferably an HIV sequence.

In a preferred embodiment, the third nucleotide sequence is resistant to cleavage directed by the gene product as a result of one or more conservative alterations in the nucleotide sequence which remove cleavage sites recognised by the at least one gene product and/or binding sites for the at least one gene product. For example, where the gene product is an EGS, the third nucleotide sequence is adapted to prevent EGS binding and/or to remove the RNase P consensus cleavage site. Alternatively, where the gene product is a ribozyme, the third nucleotide sequence is adapted to be resistant to cleavage by the ribozyme.

Preferably the third nucleotide sequence is codon optimised for expression in host cells. The host cells, which term includes producer cells and packaging cells, are typically mammalian cells.

In a particularly preferred embodiment, (i) the viral genome is an HIV genome comprising nucleotide sequences encoding anti-HIV EGSs and optionally anti-HIV ribozyme

sequences directed against HIV packaging component sequences (such as gag.pol) in a target HIV and (ii) the viral system for producing packaged HIV particles further comprises nucleotide constructs encoding the same packaging components (such as gag.pol proteins) as in the target HIV wherein the sequence of the nucleotide constructs is different from that found in the target HIV so that the anti-HIV EGS and anti-HIV ribozyme sequences cannot effect cleavage or degradation of the gag.pol transcripts during production of the HIV particles in producer cells.

The present invention also provides a viral particle comprising a viral vector according to the present invention and one or more polypeptides encoded by the third nucleotide sequences according to the present invention. For example the present invention provides a viral particle produced using the viral vector production system of the invention.

In another aspect, the present invention provides a method for producing a viral particle which method comprises introducing into a host cell (i) a viral genome vector according to the present invention; (ii) one or more third nucleotide sequences according to the present invention; and (iii) nucleotide sequences encoding the other essential viral packaging components not encoded by the one or more third nucleotide sequences.

The present invention further provides a viral particle produced using by the method of the invention.

The present invention also provides a pharmaceutical composition comprising a viral particle according to the present invention together with a pharmaceutically acceptable carrier or diluent.

The viral system of the invention or viral particles of the invention may be used to treat viral infections, particularly retroviral infections such as lentiviral infections including HIV infections. Thus the present invention provides a method of treating a viral infection which method comprises administering to a human or animal patient suffering from the viral infection an effective amount of a viral system, viral particle or pharmaceutical composition of the present invention.

The invention relates in particular to HIV-based vectors carrying anti-HIV EGSs. However, the invention can be applied to any other virus, in particular any other lentivirus, for which treatment by gene therapy may be desirable. The invention is illustrated herein for HIV, but this is not considered to limit the scope of the invention to HIV-based anti-HIV vectors.

Detailed Description of the Invention

The term "viral vector" refers to a nucleotide construct comprising a viral genome capable of being transcribed in a host cell, which genome comprises sufficient viral genetic information to allow packaging of the viral RNA genome, in the presence of packaging components, into a viral particle capable of infecting a target cell. Infection of the target cell includes reverse transcription and integration into the target cell genome, where appropriate for particular viruses. The viral vector in use typically carries heterologous coding sequences (nucleotides of interest) which are to be delivered by the vector to the target cell, for example a first nucleotide sequence encoding an EGS. A viral vector is incapable of independent replication to produce infectious viral particles within the final target cell.

The term " viral vector system" is intended to mean a kit of parts which can be used when combined with other necessary components for viral particle production to produce viral particles in host cells. For example, the first nucleotide sequence may typically be present in a plasmid vector construct suitable for cloning the first nucleotide sequence into a viral genome vector construct. When combined in a kit with a third nucleotide sequence, which will also typically be present in a separate plasmid vector construct, the resulting combination of plasmid containing the first nucleotide sequence and plasmid containing the third nucleotide sequence comprises the essential elements of the invention. Such a kit may then be used by the skilled person in the production of suitable viral vector genome constructs which when transfected into a host cell together with the plasmid containing the third nucleotide sequence, and optionally nucleic acid constructs encoding other components required for viral assembly, will lead to the production of infectious viral particles.

Alternatively, the third nucleotide sequence may be stably present within a packaging cell line that is included in the kit.

The kit may include the other components needed to produce viral particles, such as host cells and other plasmids encoding essential viral polypeptides required for viral assembly. By way of example, the kit may contain (i) a plasmid containing a first nucleotide sequence encoding an anti-HIV EGS and (ii) a plasmid containing a third nucleotide sequence encoding a modified HIV gag.pol construct which cannot be cleaved by the anti-HIV ribozyme. Optional components would then be (a) an HIV viral genome construct with suitable restriction enzyme recognition sites for cloning the first nucleotide sequence into the viral genome; (b) a plasmid encoding a VSV-G env protein. Alternatively, nucleotide sequence encoding viral polypeptides required for assembly of viral particles may be provided in the kit as packaging cell lines comprising the nucleotide sequences, for example a VSV-G expressing cell line.

The term "viral vector production system" refers to the viral vector system described above wherein the first nucleotide sequence has already been inserted into a suitable viral vector genome.

Viral vectors are typically retroviral vectors, in particular lentiviral vectors such as HIV vectors. The retroviral vector of the present invention may be derived from or may be derivable from any suitable retrovirus. A large number of different retroviruses have been identified. Examples include: murine leukemia virus (MLV), human immunodeficiency virus (HIV), simian immunodeficiency virus, human T-cell leukemia virus (HTLV). equine infectious anaemia virus (EIAV), mouse mammary tumour virus (MMTV), Rous sarcoma virus (RSV), Fujinami sarcoma virus (FuSV), Moloney murine leukemia virus (Mo-MLV), FBR murine osteosarcoma virus (FBR MSV), Moloney murine sarcoma virus (Mo-MSV), Abelson murine leukemia virus (A-MLV), Avian myelocytomatosis virus-29 (MC29), and Avian erythroblastosis virus (AEV). A detailed list of retroviruses may be found in Coffin *et al.*, 1997, "Retroviruses", Cold Spring Harbour Laboratory Press Eds: JM Coffin, SM Hughes, HE Varmus pp 758-763.

Details on the genomic structure of some retroviruses may be found in the art. By way of example, details on HIV and Mo-MLV may be found from the NCBI Genbank (Genome Accession Nos. AF033819 and AF033811, respectively).

The lentivirus group can be split even further into "primate" and "non-primate". Examples of primate lentiviruses include human immunodeficiency virus (HIV), the causative agent of human auto-immunodeficiency syndrome (AIDS), and simian immunodeficiency virus (SIV). The non-primate lentiviral group includes the prototype "slow virus" visna/maedi virus (VMV), as well as the related caprine arthritis-encephalitis virus (CAEV), equine infectious anaemia virus (EIAV) and the more recently described feline immunodeficiency virus (FIV) and bovine immunodeficiency virus (BIV).

The basic structure of a retrovirus genome is a 5' LTR and a 3' LTR, between or within which are located a packaging signal to enable the genome to be packaged, a primer binding site, integration sites to enable integration into a host cell genome and *gag, pol* and *env* genes encoding the packaging components - these are polypeptides required for the assembly of viral particles. More complex retroviruses have additional features, such as *rev* and RRE sequences in HIV, which enable the efficient export of RNA transcripts of the integrated provirus from the nucleus to the cytoplasm of an infected target cell.

In the provirus, these genes are flanked at both ends by regions called long terminal repeats (LTRs). The LTRs are responsible for proviral integration, and transcription. LTRs also serve as enhancer-promoter sequences and can control the expression of the viral genes. Encapsidation of the retroviral RNAs occurs by virtue of a *psi* sequence located at the 5' end of the viral genome.

The LTRs themselves are identical sequences that can be divided into three elements, which are called U3, R and U5. U3 is derived from the sequence unique to the 3' end of the RNA. R is derived from a sequence repeated at both ends of the RNA and U5 is derived from the sequence unique to the 5' end of the RNA. The sizes of the three elements can vary considerably among different retroviruses.

In a defective retroviral vector genome *gag*, *pol* and *env* may be absent or not functional. The R regions at both ends of the RNA are repeated sequences. U5 and U3 represent unique sequences at the 5' and 3' ends of the RNA genome respectively.

In a typical retroviral vector for use in gene therapy, at least part of one or more of the *gag*, *pol* and *env* protein coding regions essential for replication may be removed from the virus. This makes the retroviral vector replication-defective. The removed portions may even be replaced by a nucleotide sequence of interest (NOI), such as a first nucleotide sequence of the invention, to generate a virus capable of integrating its genome into a host genome but wherein the modified viral genome is unable to propagate itself due to a lack of structural proteins. When integrated in the host genome, expression of the NOI occurs - resulting in, for example, a therapeutic and/or a diagnostic effect. Thus, the transfer of an NOI into a site of interest is typically achieved by: integrating the NOI into the recombinant viral vector; packaging the modified viral vector into a virion coat; and allowing transduction of a site of interest - such as a targeted cell or a targeted cell population.

A minimal retroviral genome for use in the present invention will therefore comprise (5') R - U5 - one or more first nucleotide sequences - U3-R (3'). However, the plasmid vector used to produce the retroviral genome within a host cell/packaging cell will also include transcriptional regulatory control sequences operably linked to the retroviral genome to direct transcription of the genome in a host cell/packaging cell. These regulatory sequences may be the natural sequences associated with the transcribed retroviral sequence, i.e. the 5' U3 region, or they may be a heterologous promoter such as another viral promoter, for example the CMV promoter.

Some retroviral genomes require additional sequences for efficient virus production. For example, in the case of HIV, *rev* and RRE sequence are preferably included. However the requirement for *rev* and RRE can be reduced or eliminated by codon optimisation.

Once the retroviral vector genome is integrated into the genome of its target cell as proviral DNA, the ribozyme sequences need to be expressed. In a retrovirus, the promoter is located in the 5' LTR U3 region of the provirus. In retroviral vectors, the promoter driving expression of a therapeutic gene may be the native retroviral promoter in the 5' U3 region,

or an alternative promoter engineered into the vector. The alternative promoter may physically replace the 5' U3 promoter native to the retrovirus, or it may be incorporated at a different place within the vector genome such as between the LTRs.

Thus, the first nucleotide sequence will also be operably linked to a transcriptional regulatory control sequence to allow transcription of the first nucleotide sequence to occur in the target cell. The control sequence will typically be active in mammalian cells. The control sequence may, for example, be a viral promoter such as the natural viral promoter or a CMV promoter or it may be a mammalian promoter. It is particularly preferred to use a promoter that is preferentially active in a particular cell type or tissue type in which the virus to be treated primarily infects. Thus, in one embodiment, a tissue-specific regulatory sequences may be used. The regulatory control sequences driving expression of the one or more first nucleotide sequences may be constitutive or regulated promoters.

Replication-defective retroviral vectors are typically propagated, for example to prepare suitable titres of the retroviral vector for subsequent transduction, by using a combination of a packaging or helper cell line and the recombinant vector. That is to say, that the three packaging proteins can be provided *in trans*.

A "packaging cell line" contains one or more of the retroviral *gag, pol* and *env* genes. The packaging cell line produces the proteins required for packaging retroviral DNA but it cannot bring about encapsidation due to the lack of a *psi* region. However, when a recombinant vector carrying an NOI and a *psi* region is introduced into the packaging cell line, the helper proteins can package the *psi*-positive recombinant vector to produce the recombinant virus stock. This virus stock can be used to transduce cells to introduce the NOI into the genome of the target cells. It is preferred to use a *psi* packaging signal, called *psi* plus, that contains additional sequences spanning from upstream of the splice donor to downstream of the *gag* start codon (Bender *et al.*, 1987) since this has been shown to increase viral titres.

The recombinant virus whose genome lacks all genes required to make viral proteins can tranduce only once and cannot propagate. These viral vectors which are only capable of a single round of transduction of target cells are known as replication defective vectors.

Hence, the NOI is introduced into the host/target cell genome without the generation of potentially harmful retrovirus. A summary of the available packaging lines is presented in Coffin *et al.*, 1997 (*ibid*).

Retroviral packaging cell lines in which the *gag, pol* and *env* viral coding regions are carried on separate expression plasmids that are independently transfected into a packaging cell line are preferably used. This strategy, sometimes referred to as the three plasmid transfection method (Soneoka *et al.*, 1995) reduces the potential for production of a replication-competent virus since three recombinant events are required for wild type viral production. As recombination is greatly facilitated by homology, reducing or eliminating homology between the genomes of the vector and the helper can also be used to reduce the problem of replication-competent helper virus production.

An alternative to stably transfected packaging cell lines is to use transiently transfected cell lines. Transient transfections may advantageously be used to measure levels of vector production when vectors are being developed. In this regard, transient transfection avoids the longer time required to generate stable vector-producing cell lines and may also be used if the vector or retroviral packaging components are toxic to cells. Components typically used to generate retroviral vectors include a plasmid encoding the gag/pol proteins, a plasmid encoding the env protein and a plasmid containing an NOI. Vector production involves transient transfection of one or more of these components into cells containing the other required components. If the vector encodes toxic genes or genes that interfere with the replication of the host cell, such as inhibitors of the cell cycle or genes that induce apotosis, it may be difficult to generate stable vector-producing cell lines, but transient transfection can be used to produce the vector before the cells die. Also, cell lines have been developed using transient transfection that produce vector titre levels that are comparable to the levels obtained from stable vector-producing cell lines (Pear *et al.*, 1993).

Producer cells/packaging cells can be of any suitable cell type. Most commonly, mammalian producer cells are used but other cells, such as insect cells are not excluded. Clearly, the producer cells will need to be capable of efficiently translating the env and gag, pol mRNA. Many suitable producer/packaging cell lines are known in the art. The skilled

person is also capable of making suitable packaging cell lines by, for example stably introducing a nucleotide construct encoding a packaging component into a cell line.

As will be discussed below, where the retroviral genome encodes an inhibitory RNA molecule capable of effecting the cleavage of *gag*, *pol* and/or *env* RNA transcripts, the nucleotide sequences present in the packaging cell line, either integrated or carried on plasmids, or in the transiently transfected producer cell line, which encode gag, pol and or env proteins will be modified so as to reduce or prevent binding of the inhibitory RNA molecule(s). In this way, the inhibitory RNA molecule(s) will not prevent expression of components in packaging cell lines that are essential for packaging of viral particles.

It is highly desirable to use high-titre virus preparations in both experimental and practical applications. Techniques for increasing viral titre include using a *psi* plus packaging signal as discussed above and concentration of viral stocks. In addition, the use of different envelope proteins, such as the G protein from vesicular-stomatitis virus has improved titres following concentration to 10^9 per ml (Cosset *et al.*, 1995). However, typically the envelope protein will be chosen such that the viral particle will preferentially infect cells that are infected with the virus which it desired to treat. For example where an HIV vector is being used to treat HIV infection, the env protein used will be the HIV env protein.

Suitable first nucleotide sequences for use according to the present invention encode gene products that result in the cleavage and/or enzymatic degradation of a target nucleotide sequence, which will generally be a ribonucleotide. As particular examples, EGSs, ribozymes, and antisense sequences may be mentioned, more specifically EGSs.

External guide sequences (EGSs) are RNA sequences that bind to a complementary target sequence to form a loop in the target RNA sequence, the overall structure being a substrate for RNaseP-mediated cleavage of the target RNA sequence. The structure that forms when the EGS anneals to the target RNA is very similar to that found in a tRNA precursor. The the natural activity of RNaseP can be directed to cleave a target RNA by designing a suitable EGS. The general rules for EGS design are as follows, with reference to the generic EGSs shown in Figure 9B:

Rules for EGS design in mammalian cells (see Figure 9B)

Target sequence - All tRNA precursor molecules have a G immediately 3' of the RNaseP cleavage site (i.e. the G forms a base pair with the C at the top of the acceptor stem prior to the ACCA sequence). In addition a U is found 8 nucleotides downstream in all tRNAs. (i.e. G at position 1, U at position 8). A pyrimidine may be preferred 5' of the cut site. No other specific target sequences are required.

EGS sequence - A 7 nucleotide 'acceptor stem' analogue is optimal (5' hybridising arm). A 4 nucleotide 'D-stem' analogue is preferred (3' hybridising arm). Variation in this length may alter the reaction kinetics. This will be specific to each target site. A consensus 'T-stem and loop' analogue is essential. Minimal 5' and 3' non-pairing sequences are preferred to reduce the potential for undesired folding of the EGS RNA.

Deletion of the 'anti-codon stem and loop' analogue may be beneficial. Deletion of the variable loop can also be tolerated *in vitro* but an optimal replacement loop for the deletion of both has not been defined *in vivo*.

As with ribozymes, described below, it is preferred to use more than one EGS. Preferably, a plurality of EGSs is employed, together capable of cleaving *gag, pol* and *env* RNA of the native retrovirus at a plurality of sites. Since HIV exists as a population of quasispecies, not all of the target sequences for the EGSs will be included in all HIV variants. The problem presented by this variability can be overcome by using multiple EGs. Multiple EGSs can be included in series in a single vector and can function independently when expressed as a single RNA sequence. A single RNA containing two or more EGSs having different target recognition sites may be referred to as a multitarget EGS.

Further guidance may be obtained by reference to, for example, Werner *et al.* (1997); Werner *et al.* (1998); Ma *et al.* (1998) and Kawa *et al.* (1998).

Ribozymes are RNA enzymes which cleave RNA at specific sites. Ribozymes can be engineered so as to be specific for any chosen sequence containing a ribozyme cleavage site. Thus, ribozymes can be engineered which have chosen recognition sites in transcribed

viral sequences. By way of an example, ribozymes encoded by the first nucleotide sequence recognise and cleave essential elements of viral genomes required for the production of viral particles, such as packaging components. Thus, for retroviral genomes, such essential elements include the *gag*, *pol* and *env* gene products. A suitable ribozyme capable of recognising at least one of the gag, pol and env gene sequences, or more typically, the RNA sequences transcribed from these genes, is able to bind to and cleave such a sequence. This will reduce or prevent production of the gal, pol or env protein as appropriate and thus reduce or prevent the production of retroviral particles.

Ribozymes come in several forms, including hammerhead, hairpin and hepatitis delta antigenomic ribozymes. Preferred for use herein are hammerhead ribozymes, in part because of their relatively small size, because the sequence requirements for their target cleavage site are minimal and because they have been well characterised. The ribozymes most commonly used in research at present are hammerhead and hairpin ribozymes.

Each individual ribozyme has a motif which recognises and binds to a recognition site in the target RNA. This motif takes the form of one or more "binding arms", generally two binding arms. The binding arms in hammerhead ribozymes are the flanking sequences Helix I and Helix III, which flank Helix II. These can be of variable length, usually between 6 to 10 nucleotides each, but can be shorter or longer. The length of the flanking sequences can affect the rate of cleavage. For example, it has been found that reducing the total number of nucleotides in the flanking sequences from 20 to 12 can increase the turnover rate of the ribozyme cleaving a HIV sequence, by 10-fold (Goodchild *et al.*, 1991). A catalytic motif in the ribozyme Helix II in hammerhead ribozymes cleaves the target RNA at a site which is referred to as the cleavage site. Whether or not a ribozyme will cleave any given RNA is determined by the presence or absence of a recognition site for the ribozyme containing an appropriate cleavage site.

Each type of ribozyme recognises its own cleavage site. The hammerhead ribozyme cleavage site has the nucleotide base triplet GUX directly upstream where G is guanine, U is uracil and X is any nucleotide base. Hairpin ribozymes have a cleavage site of BCUGNYR, where B is any nucleotide base other than adenine, N is any nucleotide, Y is

cytosine or thymine and R is guanine or adenine. Cleavage by hairpin ribozymes takes places between the G and the N in the cleavage site.

The nucleic acid sequences encoding the packaging components (the "third nucleotide sequences") may be resistant to the ribozyme or ribozymes because they lack any cleavage sites for the ribozyme or ribozymes. This prohibits enzymatic activity by the ribozyme or ribozymes and therefore there is no effective recognition site for the ribozyme or ribozymes. Alternatively or additionally, the potential recognition sites may be altered in the flanking sequences which form the part of the recognition site to which the ribozyme binds. This either eliminates binding of the ribozyme motif to the recognition site, or reduces binding capability enough to destabilise any ribozyme-target complex and thus reduce the specificity and catalytic activity of the ribozyme. Where the flanking sequences only are altered, they are preferably altered such that catalytic activity of the ribozyme at the altered target sequence is negligible and is effectively eliminated.

Preferably, a series of several anti-HIV ribozymes is employed in the invention. These can be any anti-HIV ribozymes but must include one or more which cleave the RNA that is required for the expression of *gag*, *pol* or *env*. Preferably, a plurality of ribozymes is employed, together capable of cleaving *gag*, *pol* and *env* RNA of the native retrovirus at a plurality of sites. Since HIV exists as a population of quasispecies, not all of the target sequences for the ribozymes will be included in all HIV variants. The problem presented by this variability can be overcome by using multiple ribozymes. Multiple ribozymes can be included in series in a single vector and can function independently when expressed as a single RNA sequence. A single RNA containing two or more ribozymes having different target recognition sites may be referred to as a multitarget ribozyme. The placement of ribozymes in series has been demonstrated to enhance cleavage. The use of a plurality of ribozymes is not limited to treating HIV infection but may be used in relation to other viruses, retroviruses or otherwise.

Antisense technology is well known on the art. There are various mechanisms by which antisense sequences are believed to inhibit gene expression. One mechanism by which antisense sequences are believed to function is the recruitment of the cellular protein RNaseH to the target sequence/antisense construct heteroduplex which results in cleavage

and degradation of the heteroduplex. Thus the antisense construct, by contrast to ribozymes, can be said to lead indirectly to cleavage/degradation of the target sequence. Thus according to the present invention, a first nucleotide sequence may encode an antisense RNA that binds to either a gene encoding an essential/packaging component or the RNA transcribed from said gene such that expression of the gene is inhibited, for example as a result of RNaseH degradation of a resulting heteroduplex. It is not necessary for the antisense construct to encode the entire complementary sequence of the gene encoding an essential/packaging component - a portion may suffice. The skilled person will easily be able to determine how to design a suitable antisense construct.

By contrast, the nucleic acid sequences encoding the essential/packaging components of the viral particles required for the assembly of viral particles in the host cells/producer cells/packaging cells (the third nucleotide sequences) are resistant to the inhibitory RNA molecules encoded by the first nucleotide sequence. For example in the case of ribozymes, resistance is typically by virtue of alterations in the sequences which eliminate the ribozyme recognition sites. At the same time, the amino acid coding sequence for the essential/packaging components is retained so that the viral components encoded by the sequences remain the same, or at least sufficiently similar that the function of the essential/packaging components is not compromised.

The term "viral polypeptide required for the assembly of viral particles" means a polypeptide normally encoded by the viral genome to be packaged into viral particles, in the absence of which the viral genome cannot be packaged. For example, in the context of retroviruses such polypeptides would include gag, pol and env. The terms "packaging component" and "essential component" are also included within this definition.

In the case of antisense sequences, the third nucleotide sequence differs from the second nucleotide sequence encoding the target viral packaging component antisense sequence to the extent that although the antisense sequence can bind to the second nucleotide sequence, or transcript thereof, the antisense sequence can not bind effectively to the third nucleotide sequence or RNA transcribed from therefrom. The changes between the second and third nucleotide sequences will typically be conservative changes, although a small number of

amino acid changes may be tolerated provided that, as described above, the function of the essential/packaging components is not significantly impaired.

Preferably, in addition to eliminating the inhibitory RNA recognition sites, the alterations to the coding sequences for the viral components improve the sequences for codon usage in the mammalian cells or other cells which are to act as the producer cells for retroviral vector particle production. This improvement in codon usage is referred to as "codon optimisation". Many viruses, including HIV and other lentiviruses, use a large number of rare codons and by changing these to correspond to commonly used mammalian codons, increased expression of the packaging components in mammalian producer cells can be achieved. Codon usage tables are known in the art for mammalian cells, as well as for a variety of other organisms.

Thus preferably, the sequences encoding the packaging components are codon optimised. More preferably, the sequences are codon optimised in their entirety. Following codon optimisation, it is found that there are numerous sites in the wild type *gag, pol* and *env* sequences which can serve as inhibitory RNA recognition sites and which are no longer present in the sequences encoding the packaging components. In an alternative but less practical strategy, the sequences encoding the packaging components can be altered by targeted conservative alterations so as to render them resistant to selected inhibitory RNAs capable of effecting the cleavage of the wild type sequences.

An additional advantage of codon optimising HIV packaging components is that this can increase gene expression. In particular, it can render *gag, pol* expression Rev independent so that *rev* and RRE need not be included in the genome (Haas *et al.*, 1996). Rev-independent vectors are therefore possible. This in turn enables the use of anti-*rev* or RRE factors in the retroviral vector.

As described above, the packaging components for a retroviral vector include expression products of *gag, pol* and *env* genes. In accordance with the present invention, *gag* and *pol* employed in the packaging system are derived from the target retrovirus on which the vector genome is based. Thus, in the RNA transcript form, *gag* and *pol* would normally be cleavable by the ribozymes present in the vector genome. The *env* gene employed in the

packaging system may be derived from a different virus, including other retroviruses such as MLV and non-retroviruses such as VSV (a Rhabdovirus), in which case it may not need any sequence alteration to render it resistant to cleavage effected by the inhibitory RNA(s). Alternatively, *env* may be derived from the same retrovirus as *gag* and *pol*, in which case any recognition sites for the inhibitory RNA(s) will need to be eliminated by sequence alteration.

The process of producing a retroviral vector in which the envelope protein is not the native envelope of the retrovirus is known as "pseudotyping". Certain envelope proteins, such as MLV envelope protein and vesicular stomatitis virus G (VSV-G) protein, pseudotype retroviruses very well. Pseudotyping can be useful for altering the target cell range of the retrovirus. Alternatively, to maintain target cell specificity for target cells infected with the particular virus it is desired to treat, the envelope protein may be the same as that of the target virus, for example HIV.

Other therapeutic coding sequences may be present along with the first nucleotide sequence or sequences. Other therapeutic coding sequences include, but are not limited to, sequences encoding cytokines, hormones, antibodies, immunoglobulin fusion proteins, enzymes, immune co-stimulatory molecules, anti-sense RNA, a transdominant negative mutant of a target protein, a toxin, a conditional toxin, an antigen, a single chain antibody, tumour suppresser protein and growth factors. When included, such coding sequences are operatively linked to a suitable promoter, which may be the promoter driving expression of the first nucleotide sequence or a different promoter or promoters.

Thus the invention comprises two components. The first is a genome construction that will be packaged by viral packaging components and which carries a series of anti-viral inhibitory RNA molecules such as anti-HIVEGs. These could be any anti-HIV EGSs but the key issue for this invention is that some of them result in cleavage of RNA that is required for the expression of native or wild type HIV *gag*, *pol* or *env* coding sequences. The second component is the packaging system which comprises a cassette for the expression of HIV gag, pol and a cassette either for HIV env or an envelope gene encoding a pseudotyping envelope protein - the packaging system being resistant to the inhibitory RNA molecules.

The viral particles of the present invention, and the viral vector system and methods used to produce may thus be used to treat or prevent viral infections, preferably retroviral infections, in particular lentiviral, especially HIV, infections. Specifically, the viral particles of the invention, typically produced using the viral vector system of the present invention may be used to deliver inhibitory RNA molecules to a human or animal in need of treatment for a viral infection.

Alternatively, or in addition, the viral production system may be used to transfect cells obtained from a patient *ex vivo* and then returned to the patient. Patient cells transfected *ex vivo* may be formulated as a pharmaceutical composition (see below) prior to readministration to the patient.

Preferably the viral particles are combined with a pharmaceutically acceptable carrier or diluent to produce a pharmaceutical composition. Thus, the present invention also provides a pharmaceutical composition for treating an individual, wherein the composition comprises a therapeutically effective amount of the viral particle of the present invention, together with a pharmaceutically acceptable carrier, diluent, excipient or adjuvant. The pharmaceutical composition may be for human or animal usage.

The choice of pharmaceutical carrier, excipient or diluent can be selected with regard to the intended route of administration and standard pharmaceutical practice. Suitable carriers and diluents include isotonic saline solutions, for example phosphate-buffered saline. The pharmaceutical compositions may comprise as - or in addition to - the carrier, excipient or diluent any suitable binder(s), lubricant(s), suspending agent(s), coating agent(s), solubilising agent(s), and other carrier agents that may aid or increase the viral entry into the target site (such as for example a lipid delivery system).

The pharmaceutical composition may be formulated for parenteral, intramuscular, intravenous, intracranial, subcutaneous, oral, intraocular or transdermal administration.

Where appropriate, the pharmaceutical compositions can be administered by any one or more of: inhalation, in the form of a suppository or pessary, topically in the form of a

lotion, solution, cream, ointment or dusting powder, by use of a skin patch, orally in the form of tablets containing excipients such as starch or lactose, or in capsules or ovules either alone or in admixture with excipients, or in the form of elixirs, solutions or suspensions containing flavouring or colouring agents, or they can be injected parenterally, for example intracavernosally, intravenously, intramuscularly or subcutaneously. For parenteral administration, the compositions may be best used in the form of a sterile aqueous solution which may contain other substances, for example enough salts or monosaccharides to make the solution isotonic with blood. For buccal or sublingual administration the compositions may be administered in the form of tablets or lozenges which can be formulated in a conventional manner.

The amount of virus administered is typically in the range of from 10^3 to 10^{10} pfu, preferably from 10^5 to 10^8 pfu, more preferably from 10^6 to 10^7 pfu. When injected, typically 1-10 μl of virus in a pharmaceutically acceptable suitable carrier or diluent is administered.

When the polynucleotide/vector is administered as a naked nucleic acid, the amount of nucleic acid administered is typically in the range of from 1 μg to 10 mg, preferably from 100 μg to 1 mg.

Where the first nucleotide sequence (or other therapeutic sequence) is under the control of an inducible regulatory sequence, it may only be necessary to induce gene expression for the duration of the treatment. Once the condition has been treated, the inducer is removed and expression of the NOI is stopped. This will clearly have clinical advantages. Such a system may, for example, involve administering the antibiotic tetracycline, to activate gene expression via its effect on the tet repressor/VP16 fusion protein.

The invention will now be further described by way of Examples, which are meant to serve to assist one of ordinary skill in the art in carrying out the invention and are not intended in any way to limit the scope of the invention. The Examples refer to the Figures. In the Figures:

Figure 1 shows schematically ribozymes inserted into four different HIV vectors;

Figure 2 shows schematically how to create a suitable 3' LTR by PCR;

Figure 3 shows the codon usage table for wild type HIV *gag,pol* of strain HXB2 (accession number: K03455).

Figure 4 shows the codon usage table of the codon optimised sequence designated gag,pol-SYNgp.

Figure 5 shows the codon usage table of the wild type HIV *env* called env-mn.

Figure 6 shows the codon usage table of the codon optimised sequence of HIV *env* designated SYNgp160mn.

Figure 7 shows three plasmid constructs for use in the invention.

Figure 8 shows the principle behind two systems for producing retroviral vector particles.

Figure 9 A shows an EGS based on tyrosyl t-RNA

Figure 9B shows a consensus EGS sequence.

Figure 10 shows twelve different anti-HIV EGS constructs.

Figure 11 is a schematic representation of pDozenEgs and construction of pH4DozenEgs.

The invention will now be further described in the Examples which follow, which are intended as an illustration only and do not limit the scope of the invention.

EXAMPLES

Reference Example 1 - Construction of a Ribozyme-encoding Genome

The HIV *gag.pol* sequence was codon optimised (Figure 4 and SEQ I.D. No. 1) and synthesised using overlapping oligos of around 40 nucleotides. This has three advantages. Firstly it allows an HIV based vector to carry ribozymes and other therapeutic factors. Secondly the codon optimisation generates a higher vector titre due to a higher level of gene expression. Thirdly *gag.pol* expression becomes *rev* independent which allows the use of anti-*rev* or RRE factors.

Conserved sequences within *gag.pol* were identified by reference to the HIV Sequence database at Los Alamos National Laboratory (http:// hiv-web.lanl.gov/) and used to design ribozymes. Because of the variability between subtypes of HIV-1 the ribozymes were designed to cleave the predominant subtype within North America, Latin America and the Caribbean, Europe, Japan and Australia; that is subtype B. The sites chosen were cross-referenced with the synthetic *gagpol* sequence to ensure that there was a low possibility of cutting the codon optimised gagpol mRNA. The ribozymes were designed with *Xho*I and *Sal*I sites at the 5' and 3' end respectively. This allows the construction of separate and tandem ribozymes.

The ribozymes are hammerhead (Riddell *et al.*, 1996) structures of the following general structure:

Helix I	Helix II	Helix III
5 ' - NNNNNNNN~	CUGAUGAGGCCGAAAGGCCGAA	~NNNNNNNN~

The catalytic domain of the ribozyme (Helix II) can tolerate some changes without reducing catalytic turnover.

The cleavage sites, targeting *gag* and *pol*, with the essential GUX triplet (where X is any nucleotide base) are as follows:

GAG 1	5 '	UAGUAAGAAUGUAUAGCCCUAC
GAG 2	5 '	AACCCAGAUUGUAAGACUAUUU
GAG 3	5 '	UGUUUCAAUUGUGGCAAAGAAG
GAG 4	5 '	AAAAAGGGCUGUUGGAAAUGUG
POL 1	5 '	ACGACCCCUCGUCACAAUAAAG
POL 2	5 '	GGAAUUGGAGGUUUUAUCAAAG
POL 3	5 '	AUAUUUUUCAGUUCCCUUAGAU
POL 4	5 '	UGGAUGAUUUGUAUGUAGGAUC
POL 5	5 '	CUUUGGAUGGGUUAUGAACUCC
POL 6	5 '	CAGCUGGACUGUCAAUGACAUA
POL 7	5 '	AACUUUCUAUGUAGAUGGGGCA
POL 8	5 '	AAGGCCGCCUGUUGGUGGGCAG
POL 9	5 '	UAAGACAGCAGUACAAAUGGCA

The ribozymes are inserted into four different HIV vectors (pH4 (Gervaix *et al.*, 1997), pH6, pH4.1, or pH6.1) (Figure 1). In pH4 and pH6, transcription of the ribozymes is driven by an internal HCMV promoter (Foecking *et al.*, 1986). From pH4.1 and pH6.1, the ribozymes are expressed from the 5' LTR. The major difference between pH4 and pH6 (and pH4.1 and pH6.1) resides in the 3' LTR in the production plasmid. pH4 and pH4.1 have the HIV U3 in the 3' LTR. pH6 and pH6.1 have HCMV in the 3'LTR. The HCMV promoter replaces most of the U3 and will drive expression at high constitutive levels while the HIV-1 U3 will support a high level of expression only in the presence of Tat.

The HCMV/HIV-1 hybrid 3' LTR is created by recombinant PCR with three PCR primers (Figure 2). The first round of PCR is performed with RIB1 and RIB2 using pH4 (Kim *et al.*, 1998) as the template to amplify the HIV-1 HXB2 sequence 8900-9123. The second round of PCR makes the junction between the 5' end of the HIV-1 U3 and the HCMV promoter by amplifying the hybrid 5' LTR from pH4. The PCR product from the first PCR reaction and RIB3 serves as the 5' primer and 3' primer respectively.

RIB1: 5′-CAGCTGCTCGAGCAGCTGAAGCTTGCATGC-3′

RIB2: 5′-GTAAGTTATGTAACGGACGATATCTTGTCTTCTT-3′

RIB3: 5′-CGCATAGTCGACGGGCCCGCCACTGCTAGAGATTTTC-3′

The PCR product is then cut with *Sph*I and *Sal*I and inserted into pH4 thereby replacing the 3' LTR. The resulting plasmid is designated pH6. To construct pH4.1 and pH6.1, the internal HCMV promoter *(Spe*I - *Xho*I*)* in pH4 and pH6 is replaced with the polycloning site of pBluescript II KS+ (Stratagene) *(Spe*I - *Xho*I*)*.

The ribozymes are inserted into the *Xho*I sites in the genome vector backbones. Any ribozymes in any configuration could be used in a similar way.

Reference Example 2 - Construction of a Packaging System

The packaging system can take various forms. In a first form of packaging system, the HIV gag, pol components are co-expressed with the HIV env coding sequence. In this case, both the gag, pol and the env coding sequences are altered such that they are resistant to the anti-HIV ribozymes that are built into the genome. At the same time as altering the codon usage to achieve resistance, the codons can be chosen to match the usage pattern of the most highly expressed mammalian genes. This dramatically increases expression levels and so increases titre. A codon optimised HIV env coding sequence has been described by Haas *et al.* (1996). In the present example, a modified codon optimised HIV env sequence is used (SEQ I.D. No. 3). The corresponding env expression plasmid is designated pSYNgp160mn. The modified sequence contains extra motifs not used by Haas *et al.* The extra sequences were taken from the HIV env sequence of strain MN and codon optimised. Any similar modification of the nucleic acid sequence would function similarly as long as it used codons corresponding to abundant tRNAs (Zolotukhin *et al.*, 1996) and lead to resistance to the ribozymes in the genome.

In one example of a gag, pol coding sequence with optimised codon usage, overlapping oligonucleotides are synthesised and then ligated together to produce the synthetic coding sequence. The sequence of a wild-type (Genbank accession no. K03455) and synthetic (gagpol-SYNgp) gagpol sequence is shown in SEQ I.D. Nos 1 and 2, respectively and their codon usage is shown in Figures 3 and 4, respectively. The sequence of a wild type env coding sequence (Genbank Accession No. M17449) is given in SEQ I.D. No 3, the sequence of a synthetic codon optimised sequence is given in SEQ. I.D. No. 4 and their

codon usage tables are given in Figures 5 and 6, respectively. As with the env coding sequence any gag, pol sequence that achieves resistance to the ribozymes could be used. The synthetic sequence shown is designated gag, pol-SYNgp and has an *EcoR*I site at the 5' end and a *Not*l site at the 3' end. It is inserted into pClneo (Promega) to produce plasmid pSYNgp.

The sequence of the codon optimised gagpol sequence is shown in SEQ I.D. No. 2. This sequence starts at the ATG and ends at the stop codon of gagpol. The wild type sequence is retained around the frameshift site so that the right amount of gagpol is made.

In addition other constructs can be used that contain the optimised gagpol of pSYNgp but also have differing amounts of the wild type HIV 1 sequence of strain HXB2 (accession number: K03455) at the 5' end. These constructs are described below (the start ATG of pSYNgp is shown in bold in these sequences).

pSYNgp2 contains the entire leader sequence of HIV-1 (SEQ ID. No. 12).
pSYNgp3 contains the leader sequence of HIV-1 from the major splice donor (SEQ ID. No. 13).
pSYNgp4 contains 20pb of the leader sequence of HIV-1 upstream of the start codon of ATG (SEQ ID. No. 14).

These constructs may be made by overlapping PCR. Using appropriate restriction enzymes these sequences can be inserted into mammalian expression vectors such as pCI-Neo (Promega). All these gag/pol constructs can be used to supply HIV gag/pol for the generation of viral vectors. These viral vectors can be used to express either EGS molecules or ribozyme molecules or antisense molecules or any peptides or proteins.

In a second form of the packaging system a synthetic gag, pol cassette is coexpressed with a non-HIV envelope coding sequence that produces a surface protein that pseudotypes HIV. This could be for example VSV-G (Ory *et al.*, 1996; Zhu *et al.*, 1990), amphotropic MLV env (Chesebro *et al.*, 1990; Spector *et al.*, 1990) or any other protein that would be incorporated into the HIV particle (Valsesia-Wittman, 1994). This includes molecules capable of targeting the vector to specific tissues. Coding sequences for non-HIV envelope

proteins not cleaved by the ribozymes and so no sequence modification is required (although some sequence modification may be desirable for other reasons such as optimisation for codon usage in mammalian cells).

Reference Example 3 - Vector Particle Production

Vector particles can be produced either from a transient three-plasmid transfection system similar to that described by Soneoka *et al.* (1995) or from producer cell lines similar to those used for other retroviral vectors (Ory *et al.*, 1996; Srinivasakumar *et al.*, 1997; Yu *et al.*, 1996). These principles are illustrated in Figures 7 and 8. For example, by using pH6Rz, pSYNgp and pRV67 (VSV-G expression plasmid) in a three plasmid transfection of 293T cells (Figure 8), as described by Soneoka *et al.* (1995), vector particles designated H6Rz-VSV are produced. These transduce the H6Rz genome to CD4+ cells such as C1866 or Jurkat and produce the multitarget ribozymes. HIV replication in these cells is now severely restricted.

Example 1 - Use of external guide sequences for inhibiting HIV

Ribonuclease P is a nuclear localised enzyme consisting of protein and RNA subunits. It has been found in all organisms examined and is one of the most abundant, stable and efficient enzymes in cells. Its enzymatic activity is responsible for the maturation of the 5' termini of all tRNAs which account for about 2% of the total cellular RNA.

For tRNA processing, it has been shown that RNAse P recognises a secondary structure of the tRNA. However extensive studies have shown that any complex of two RNA molecules which resemble the one tRNA molecule will also be recognised and cleaved by RNase P. Consequently the natural activity of RNase P can and has been successfully re-directed to target other RNA species (see Yaun and Altman, 1994, and references therein). This is achieved by engineering a sequence, containing the flanking motif recognised by RNaseP, to bind the desired target sequence. These sequences are called external guide sequence (EGSs).

Outlined here is a strategy employing the EGS system against HIV RNA. Shown in Figure 2 A, B and C are twelve EGS sequences designed to target twelve separate HIV gag/pol sequences. These target sequences are conserved throughout the clade B of HIV. The sequence numbering in each figure designates the position of the required conserved G of each target sequences based on the HXB2 published sequence.

The external guide sequences shown here all have anticodon stem-loops deleted. These are non-limiting examples; for instance full length 3/4 tRNA based EGSs might be used if preferred (see Yuan and Altman, 1994).

Outlined in SEQ ID. Nos. 5 to 10 (see below) and Figure 11 is the cloning strategy employed to construct an HIV vector containing the EGSs described in SEQ ID. Nos. 5 to 10. The oligonucleotides prefixed 1, 2, 3, 4, 5 and 6 are respectively annealed together and sequentially cloned into the pSP72 (Promega) cloning vector starting with the oligo. duplex 1/1A being cloned into the *Xho*I-*Sal*I site such that the EGS 4762 and EGS 4715 are orientated away from the ampicillin gene. The remaining oligonucleotides (with *Xho*I ends) are subsequently cloned stepwise (starting with oligo. duplex 2/2A, ending with duplex 6/6A) into the unique *Sal*I site (present within the terminus of the each preceding oligonucleotide) to create the plasmid pDOZENEGS. The EGSs from this vector are then transferred by *Xho*I-*Sph*I digest into the pH4Z similarily cut such that the multiple EGSs cassette replaces the lacZ gene of pH4Z (Kim *et al.*, 1998). The resulting vector is named pH4DOZENEGS (see SEQ ID. No. 11 for complete sequence).

Egs 1/1A (SEQ ID. No. 5)

```
    XhoI
5'- tcgagcccggggatgacgtcatcgacttcgaaggttcgaatccttctactgccaccatttttt
        cgggcccctactgcagtagctgaagcttccaagcttaggaagatgacggtggtaaaaaa
ctctacgtcatcgacttcgaaggttcgaatccttccctgtccaccagtcgacc-3'
gagatgcagtagctgaagcttccaagcttaggaagggacaggtggtcagctggagct-5'
```

Egs 2/2A (SEQ ID. No. 6)

```
5'- tcgagtattacgtcatcgacttcgaaggttcgaatccttctagattcaccatttttttaggaacg
        cataatgcagtagctgaagcttccaagcttaggaagtactaagtggtaaaaaatccttgc
```

tcatcgacttcgaaggttcgaatccttc**cagttcc**accagtcgacc-3′
agtagctgaagcttccaagcttagga**aggtcaagg**tggtcagctggagct-5′

Egs 3/3A (SEQ ID. No. 7)

5′- tcga**ggccaa**cgtcatcgacttcgaaggttcgaatccttc**tcttccc**accatttttt**ttcc**
ccggttgcagtagctgaagcttccaagcttaggaag**agaagggt**ggtaaaaaa**aagg**

acgtcatcgacttcgaaggttcgaatccttc**ggggccc**accagtcgacc-3′
tgcagtagctgaagcttccaagcttaggaag**ccccggg**tggtcagctggagct-5′

Egs 4/4A (SEQ ID. No. 8)

5′- tcga**gggct**acgtcatcgacttcgaaggttcgaatccttc**ttgcttc**accatttttt
cccgatgcagtagctgaagcttccaagcttaggaag**aacgaagt**ggtaaaaaa

ctgaacgtcatcgacttcgaaggttcgaatccttc**tgctgtc**accagtcgacc-3′
gacttgcagtagctgaagcttccaagcttaggaag**acgacag**tggtcagctggagct-5′

Egs 5/5A (SEQ ID. No. 9)

5′- tcga**gtataa**cgtcatcgacttcgaaggttcgaatccttc**accggtc**accatttttt**tata**
catattgcagtagctgaagcttccaagcttaggaag**tggccagt**ggtaaaaaaa**tat**

acgtcatcgacttcgaaggttcgaatccttc**ttcttac**accagtcgacc-3′
tgcagtagctgaagcttccaagcttaggaag**aagaatg**tggtcagctggagct-5′

Egs 6/6A (SEQ ID. No. 10)

5′- tcga**ggtac**acgtcatcgacttcgaaggttcgaatccttc**gtagttc**accatttttt**gtgc**
ccatgtgcagtagctgaagcttccaagcttaggaag**catcaagt**ggtaaaaaa**cacg**

SphI
acgtcatcgacttcgaaggttcgaatccttc**taggccc**accagtcgacgcatgcc-3′
tgcagtagctgaagcttccaagcttaggaag**atccggg**tggtcagctgcgtacggagct-5′

The pH4DOZENEGS_vector may be used to both deliver and express the example EGS sequences to appropriate eukaryotic cells in a manner as described for ribozymes in reference examples 1, 2 and 3 whereby the use of a codon optimised gag/pol and env genes would prevent EGSs from targeting these genes during viral production. The inclusion of the EGS sequences into an HIV derived vector will not only allow expression of such sequences in the target cell but also packaging and transfer of such therapeutic sequences by the patient's own HIV. These example EGS sequences target HIV RNA for cleavage by RNAse P. This example is not limiting and other suitable EGS and derived sequences may also be used; be they expressed singularly, in multiples, from pol I, pol II or pol III promoters and derivatives thereof and/or in combination with other HIV treatments. Other

appropriate nucleotide sequences of interest (NOIs) may also be included in combination with EGSs if preferred.

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

References

Bender *et al.*, 1987, J Virol 61: 1639-1646.

Chesebro, B., K. Wehrly, and W. Maury. 1990. J Virol. 64:4553-7.

Cosset *et al.*, 1995, J. Virol. 69: 7430-7436.

Foecking, M. K., and H. Hofstetter. 1986. Gene. 45:101-105.

Forster and Altman, 1990, Science 249: 783-786.

Gervaix, A., X. Li, G. Kraus, and F. Wong Staal. 199. J Virol. 71 :3048-53.

Goodchild, J., V. Kohli. 1991. *Arch Biochem Biophys* Feb 1; 284(2):386-391.

Haas, J., E.-C. Park, and B. Seed. 1996. Current Biology. 6:315.

Kawa *et al.*, 1998, RNA 4: 1397-1406.

Kim, V. N., K. Mitrophanous, S. M. Kingsman, and K. A. J. 1998. J Virol 72: 811-816.

Lever, A. M. 1995. Br Med Bull. 51:149-66.

Ma *et al.*, 1998, Antisense and Nucleic Acid Drug Development 8: 415-426.

Ory, D. S., B. A. Neugeboren, and R. C. Mulligan. 1996. Proc Natl Acad Sci U S A. 93:11400-6.

Pear *et al.*, 1993, Proc Natl Acad Sci 90: 8392-8396.

Riddell, S. R., M. Elliott, D. A. Lewinsohn, M. J. Gilbert, L. Wilson, S. A. Manley, S. D. Lupton, R. W. Overell, T. C. Reynolds, L. Corey, and P. D. Greenberg. 1996. Nat Med. 2:216-23.

Soneoka, Y., P. M. Cannon, E. E. Ramsdale, J. C. Griffiths, G. Romano, S. M. Kingsman, and A. J. Kingsman. 1995. Nucleic Acids Res. 23:628-33.

Spector, D. H., E. Wade, D. A. Wright, V. Koval, C. Clark, D. Jaquish, and S. A. Spector. 1990. J Virol. 64:2298-2308.

Srinivasakumar, N., N. Chazal, C. Helga Maria, S. Prasad, M. L. Hammarskjold, and D. Rekosh. 1997. J Virol. 71 :5841-8.

Valsesia Wittmann, S., A. Drynda, G. Deleage, M. Aumailley, J. M. Heard, O. Danos, G. Verdier, and F. L. Cosset. 1994. J Virol. 68:4609-19.

Werner *et al.*, 1997, Nucleic Acids Symposium Series No. 36: 19-21.

Werner *et al.*, 1998, RNA 4: 847-855.

Yu, H., A. B. Rabson, M. Kaul, Y. Ron, and J. P. Dougherty. 1996. J Virol. 70:4530-37.

Yuan and Altman, 1994, Science 263:1269-1273.

Yuan and Altman, 1995, EMBO J. 14: 159-168.

Yuan *et al.*, 1992, Proc Natl Acad Sci 89: 8006-8010.

Zhu, Z. H., S. S. Chen, and A. S. Huang. 1990. J Acquir Immune Defic Syndr. 3:215-9.

Zolotukhin, S., M. Potter, W. W. Hauswirth, J. Guy, and N. Muzyczka. 1996. J Virol. 70:4646-54.

CLAIMS

1. A viral vector system comprising:

(i) a first nucleotide sequence encoding an external guide sequence capable of binding to and effecting the cleavage by RNase P of a second nucleotide sequence, or transcription product thereof, encoding a viral polypeptide required for the assembly of viral particles; and

(ii) a third nucleotide sequence encoding said viral polypeptide required for the assembly of viral particles, which third nucleotide sequence has a different nucleotide sequence to the second nucleotide sequence such that the third nucleotide sequence, or transcription product thereof, is resistant to cleavage directed by the external guide sequence.

2. A system according to claim 1 further comprising at least one further first nucleotide sequence encoding a gene product capable of binding to and effecting the cleavage, directly or indirectly, of a second nucleotide sequence, or transcription product thereof, encoding a viral polypeptide required for the assembly of viral particles, wherein the gene product is selected from an external guide sequence, a ribozyme and an anti-sense ribonucleic acid.

3. A viral vector production system comprising:

(i) a viral genome comprising at least one first nucleotide sequence encoding a gene product capable of binding to and effecting the cleavage, directly or indirectly, of a second nucleotide sequence, or transcription product thereof, encoding a viral polypeptide required for the assembly of viral particles;

(ii) a third nucleotide sequence encoding said viral polypeptide required for the assembly of the viral genome into viral particles, which third nucleotide sequence has a different nucleotide sequence to the second nucleotide sequence such that said third nucleotide sequence, or transcription product thereof, is resistant to cleavage directed by said gene product;

wherein at least one of the gene products is an external guide sequence capable of binding to and effecting the cleavage by RNase P of the second nucleotide sequence.

4. A system according to claim 3 wherein in addition to an external guide sequence, at least one gene product is selected from a ribozyme and an anti-sense ribonucleic acid.

5. A system according to any one of claims 1 to 4 wherein the viral vector is a retroviral vector.

6. A system according to claim 5 wherein the retroviral vector is a lentiviral vector.

7. A system according to claim 6 wherein the lentiviral vector is an HIV vector.

8. A system according to any one of claims 5 to 7 wherein the polypeptide required for the assembly of viral particles is selected from gag, pol and env proteins.

9. A system according to claim 8 wherein at least the gag and pol proteins are from a lentivirus.

10. A system according to claim 7 wherein the env protein is from a lentivirus.

11. A system according to claim 9 or 10 wherein the lentivirus is HIV.

12. A system according to any one of the preceding claims wherein the third nucleotide sequence is resistant to cleavage directed by the gene product as a result of one or more conservative alterations in the nucleotide sequence which remove cleavage sites recognised by the at least one gene product and/or binding sites for the at least one gene product

13. A system according to any one of claims 1 to 11 wherein the third nucleotide sequence is adapted to be resistant to cleavage by the at least one gene product.

14. A system according to any one of the preceding claims wherein the third nucleotide sequence is codon optimised for expression in producer cells.

15. A system according to claim 14, wherein the producer cells are mammalian cells.

16. A system according to any one of the preceding claims comprising a plurality of first nucleotide sequences and third nucleotide sequences as defined therein.

17. A viral particle comprising a viral vector genome as defined in any one of claims 3 to 16 and one or more third nucleotide sequences as defined in any of claims 3 to 16.

18. A viral particle produced using a viral vector production system according to any one of claims 3 to 16.

19. A method for producing a viral particle which method comprises introducing into a host cell (i) a viral genome as defined in any one of claims 3 to 16 (ii) one or more third nucleotide sequences as defined in any of claims 3 to 16 and (iii) nucleotide sequences encoding the other essential viral packaging components not encoded by the one or more third nucleotide sequences.

20. A viral particle produced by the method of claim 19.

21. A pharmaceutical composition comprising a viral particle according to claims 17, 18 or 20 together with a pharmaceutically acceptable carrier or diluent.

22. A viral system according to any one of claims 1 to 17 or a viral particle according to claims 17, 18 or 20 in treating a viral infection.

23. A viral system according to any one of claims 1 to 17 for use in a method of producing viral particles.

Figure 1

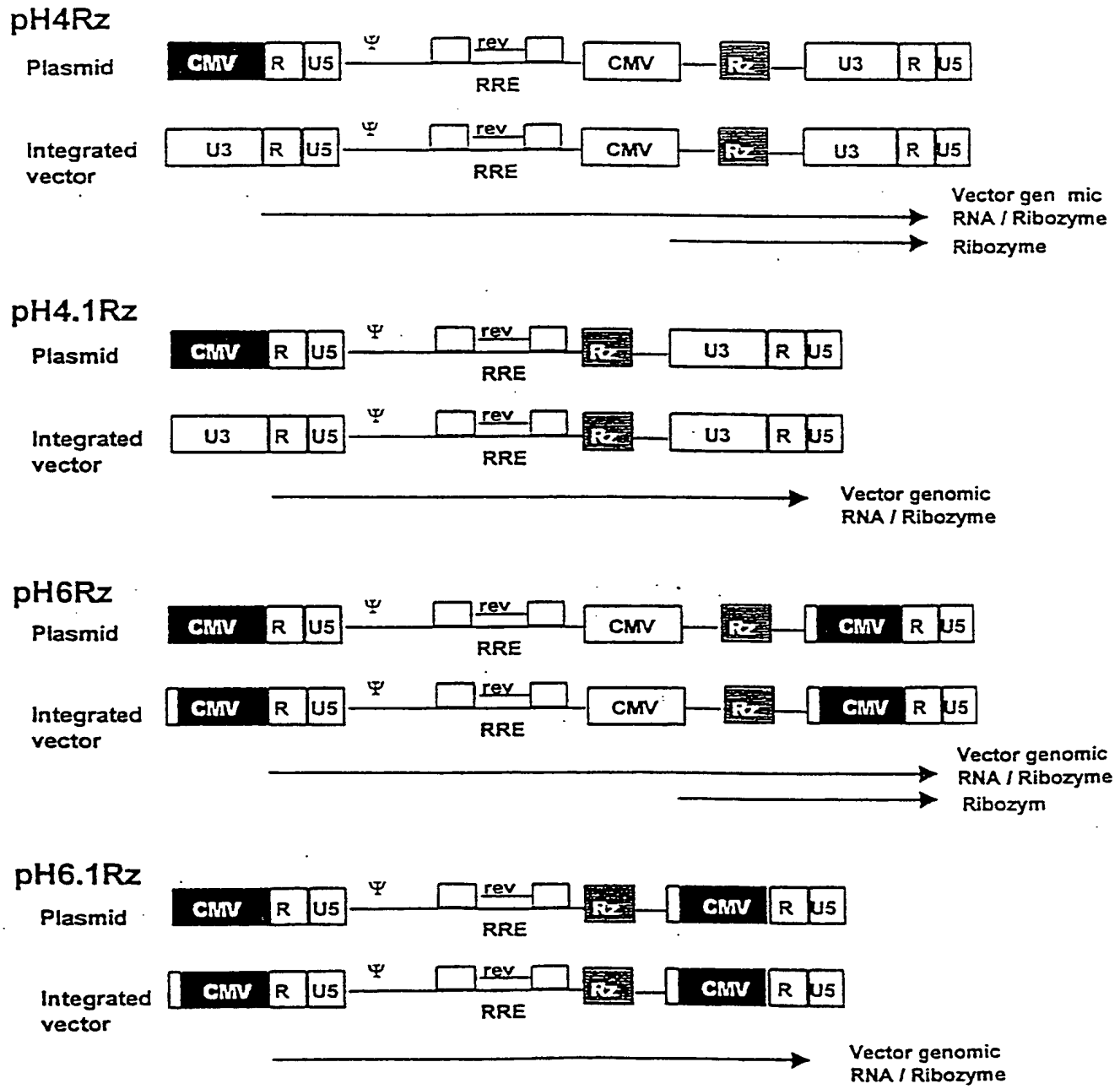

pH4Rz
Plasmid
CMV
R
U5
Ψ
rev
RRE
CMV
Rz
U3
R
U5
Integrated vector
U3
R
U5
Ψ
rev
RRE
CMV
Rz
U3
R
U5
Vector gen mic RNA / Ribozyme
Ribozyme
pH4.1Rz
Plasmid
CMV
R
U5
Ψ
rev
RRE
Rz
U3
R
U5
Integrated vector
U3
R
U5
Ψ
rev
RRE
Rz
U3
R
U5
Vector genomic RNA / Ribozyme
pH6Rz
Plasmid
CMV
R
U5
Ψ
rev
RRE
CMV
Rz
CMV
R
U5
Integrated vector
CMV
R
U5
Ψ
rev
RRE
CMV
Rz
CMV
R
U5
Vector genomic RNA / Ribozyme
Ribozym
pH6.1Rz
Plasmid
CMV
R
U5
Ψ
rev
RRE
Rz
CMV
R
U5
Integrated vector
CMV
R
U5
Ψ
rev
RRE
Rz
CMV
R
U5
Vector genomic RNA / Ribozyme

figure 2

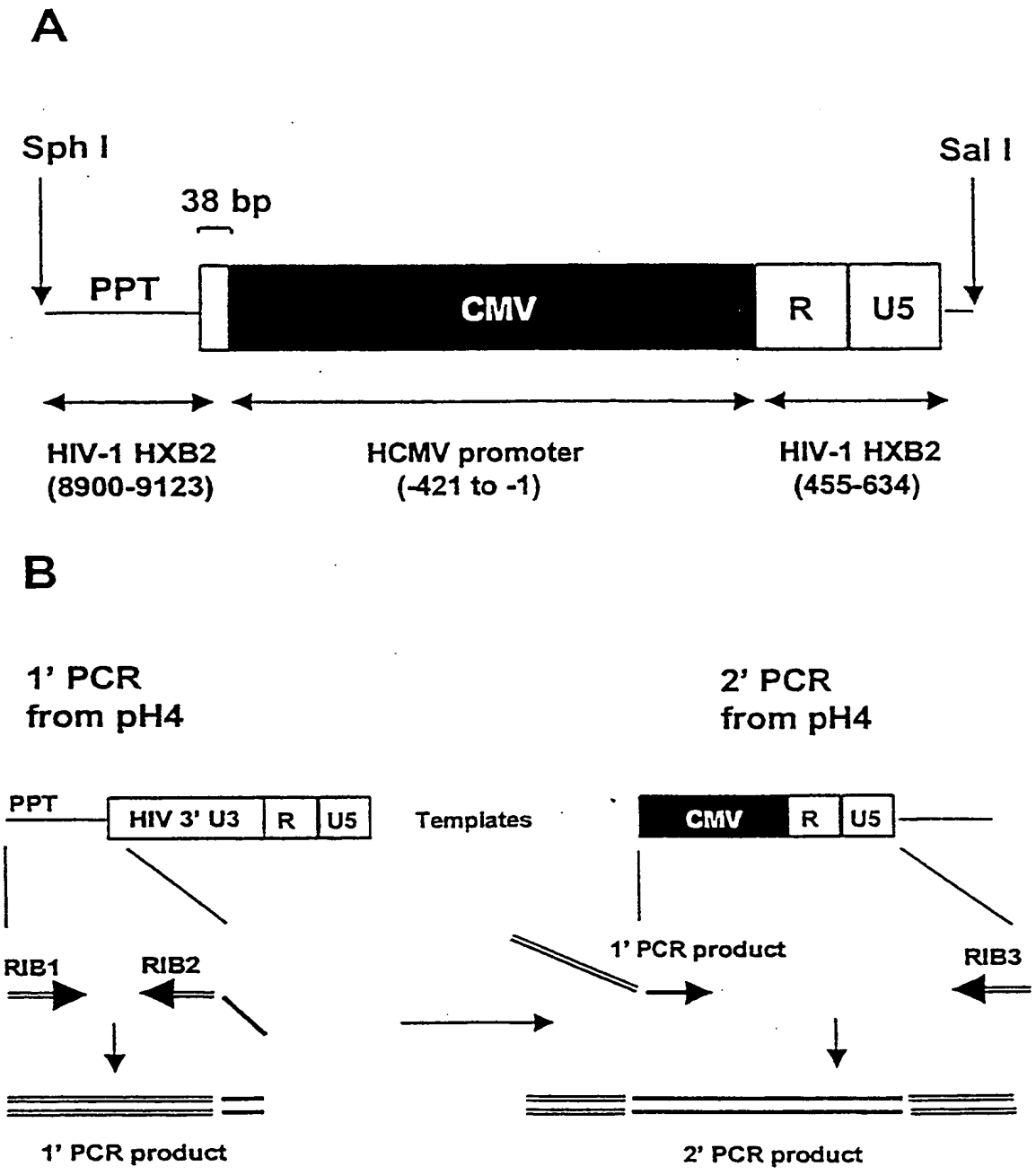

A
Sph I
Sal I
38 bp
PPT
CMV
R
U5
HIV-1 HXB2
(8900-9123)
HCMV promoter
(-421 to -1)
HIV-1 HXB2
(455-634)
B
1' PCR
from pH4
2' PCR
from pH4
PPT
HIV 3' U3
R
U5
Templates
CMV
R
U5
RIB1
RIB2
1' PCR product
RIB3
1' PCR product
2' PCR product

Figure 3

gagpol-HXB2 -> Codon Usage

DNA sequence 4308 b.p. ATGGGTGCGAGA ... GATGAGGATTAG linear

1436 codons

MW : 161929 Dalton CAI(S.c.) : 0.083 CAI(E.c.) : 0.151

TTT	phe	F	21	TCT	ser	S	3	TAT	tyr	Y	30
TTC	phe	F	14	TCC	ser	S	3	TAC	tyr	Y	9
TTA	leu	L	46	TCA	ser	S	19	TAA	OCH	Z	-
TTG	leu	L	11	TCG	ser	S	1	TAG	AMB	Z	1
CTT	leu	L	13	CCT	pro	P	21	CAT	his	H	20
CTC	leu	L	7	CCC	pro	P	14	CAC	his	H	7
CTA	leu	L	17	CCA	pro	P	41	CAA	gln	Q	56
CTG	leu	L	16	CCG	pro	P	-	CAG	gln	Q	39
ATT	ile	I	30	ACT	thr	T	24	AAT	asn	N	42
ATC	ile	I	14	ACC	thr	T	20	AAC	asn	N	16
ATA	ile	I	56	ACA	thr	T	43	AAA	lys	K	88
ATG	met	M	29	ACG	thr	T	1	AAG	lys	K	34
GTT	val	V	15	GCT	ala	A	17	GAT	asp	D	37
GTC	val	V	11	GCC	ala	A	19	GAC	asp	D	26
GTA	val	V	55	GCA	ala	A	55	GAA	glu	E	75
GTG	val	V	15	GCG	ala	A	5	GAG	glu	E	32

TGT	cys	C	18								
TGC	cys	C	2								
TGA	OPA	Z	-								
TGG	trp	W	37								
CGT	arg	R	-								
CGC	arg	R	-								
CGA	arg	R	3								
CGG	arg	R	3								
AGT	ser	S	18								
AGC	ser	S	16								
AGA	arg	R	45								
AGG	arg	R	18								
GGT	gly	G	11								
GGC	gly	G	10								
GGA	gly	G	61								
GGG	gly	G	26								

Figure 4

gagpol-SYNgp [1 to 4308] -> Codon Usage

DNA sequence 4308 b.p. ATGGGCGCCCGC ... GATGAGGATTAG linear

1436 codons

MW : 161929 Dalton CAI(S.c.) : 0.080 CAI(E.c.) : 0.296

TTT	phe	F	5	TCT	ser	S	5	TAT	tyr	Y	10
TTC	phe	F	30	TCC	ser	S	11	TAC	tyr	Y	29
TTA	leu	L	2	TCA	ser	S	4	TAA	OCH	Z	-
TTG	leu	L	7	TCG	ser	S	6	TAG	AMB	Z	1
CTT	leu	L	3	CCT	pro	P	14	CAT	his	H	6
CTC	leu	L	22	CCC	pro	P	39	CAC	his	H	21
CTA	leu	L	6	CCA	pro	P	10	CAA	gln	Q	14
CTG	leu	L	70	CCG	pro	P	13	CAG	gln	Q	81
ATT	ile	I	17	ACT	thr	T	11	AAT	asn	N	13
ATC	ile	I	79	ACC	thr	T	48	AAC	asn	N	45
ATA	ile	I	4	ACA	thr	T	13	AAA	lys	K	25
ATG	met	M	29	ACG	thr	T	16	AAG	lys	K	97
GTT	val	V	5	GCT	ala	A	15	GAT	asp	D	19
GTC	val	V	27	GCC	ala	A	56	GAC	asp	D	44
GTA	val	V	6	GCA	ala	A	13	GAA	glu	E	29
GTG	val	V	58	GCG	ala	A	12	GAG	glu	E	78

TGT	cys	C	6								
TGC	cys	C	14								
TGA	OPA	Z	-								
TGG	trp	W	37								
CGT	arg	R	2								
CGC	arg	R	34								
CGA	arg	R	3								
CGG	arg	R	10								
AGT	ser	S	7								
AGC	ser	S	27								
AGA	arg	R	7								
AGG	arg	R	13								
GGT	gly	G	10								
GGC	gly	G	54								
GGA	gly	G	16								
GGG	gly	G	28								

Figure 5

env-mn [1 to 2571] -> Codon Usage

DNA sequence 2571 b.p. ATGAGAGTGAAG ... GCTTTGCTATAA linear

857 codons

MW : 97078 Dalton CAI(S.c.) : 0.083 CAI(E.c.) : 0.140

TTT	phe	F	13	TCT	ser	S	7	TAT	tyr	Y	15
TTC	phe	F	11	TCC	ser	S	3	TAC	tyr	Y	7
TTA	leu	L	20	TCA	ser	S	13	TAA	OCH	Z	1
TTG	leu	L	17	TCG	ser	S	2	TAG	AMB	Z	-
CTT	leu	L	9	CCT	pro	P	5	CAT	his	H	8
CTC	leu	L	11	CCC	pro	P	9	CAC	his	H	6
CTA	leu	L	12	CCA	pro	P	12	CAA	gln	Q	22
CTG	leu	L	15	CCG	pro	P	2	CAG	gln	Q	19
ATT	ile	I	21	ACT	thr	T	16	AAT	asn	N	50
ATC	ile	I	10	ACC	thr	T	14	AAC	asn	N	13
ATA	ile	I	32	ACA	thr	T	28	AAA	lys	K	32
ATG	met	M	17	ACG	thr	T	5	AAG	lys	K	14
GTT	val	V	8	GCT	ala	A	16	GAT	asp	D	18
GTC	val	V	9	GCC	ala	A	7	GAC	asp	D	14
GTA	val	V	26	GCA	ala	A	20	GAA	glu	E	36
GTG	val	V	12	GCG	ala	A	5	GAG	glu	E	10

TGT	cys	C	16								
TGC	cys	C	5								
TGA	OPA	Z	-								
TGG	trp	W	30								
CGT	arg	R	-								
CGC	arg	R	2								
CGA	arg	R	1								
CGG	arg	R	1								
AGT	ser	S	18								
AGC	ser	S	11								
AGA	arg	R	30								
AGG	arg	R	15								
GGT	gly	G	10								
GGC	gly	G	6								
GGA	gly	G	28								
GGG	gly	G	12								

Figure 6

SYNgp160mn -> Codon Usage

DNA sequence 2571 b.p. ATGAGGGTGAAG ... GCGCTGCTGTAA linear

857 codons

MW : 97078 Dalton CAI(S.c.) : 0.074 CAI(E.c.) : 0.419

TTT	phe	F	-	TCT	ser	S	2	TAT	tyr	Y	1	TGT	cys	C	-
TTC	phe	F	24	TCC	ser	S	4	TAC	tyr	Y	21	TGC	cys	C	21
TTA	leu	L	-	TCA	ser	S	-	TAA	OCH	Z	1	TGA	OPA	Z	-
TTG	leu	L	-	TCG	ser	S	-	TAG	AMB	Z	-	TGG	trp	W	30
CTT	leu	L	-	CCT	pro	P	-	CAT	his	H	2	CGT	arg	R	1
CTC	leu	L	20	CCC	pro	P	26	CAC	his	H	12	CGC	arg	R	36
CTA	leu	L	1	CCA	pro	P	-	CAA	gln	Q	-	CGA	arg	R	-
CTG	leu	L	63	CCG	pro	P	2	CAG	gln	Q	41	CGG	arg	R	4
ATT	ile	I	2	ACT	thr	T	-	AAT	asn	N	2	AGT	ser	S	-
ATC	ile	I	61	ACC	thr	T	59	AAC	asn	N	61	AGC	ser	S	48
ATA	ile	I	-	ACA	thr	T	-	AAA	lys	K	1	AGA	arg	R	2
ATG	met	M	17	ACG	thr	T	4	AAG	lys	K	45	AGG	arg	R	6
GTT	val	V	-	GCT	ala	A	-	GAT	asp	D	2	GGT	gly	G	1
GTC	val	V	1	GCC	ala	A	40	GAC	asp	D	30	GGC	gly	G	47
GTA	val	V	1	GCA	ala	A	-	GAA	glu	E	3	GGA	gly	G	-
GTG	val	V	53	GCG	ala	A	8	GAG	glu	E	43	GGG	gly	G	8

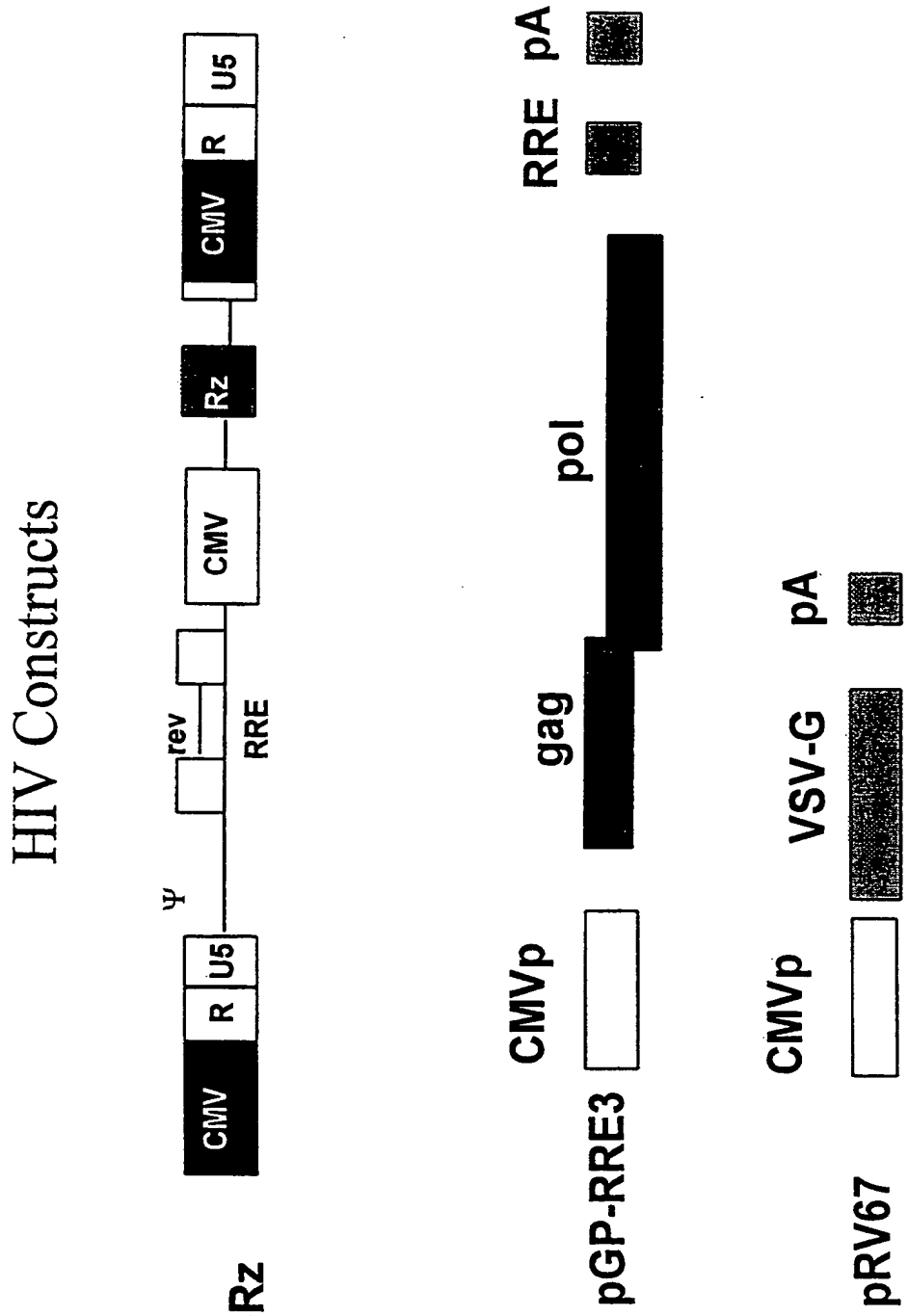

HIV Constructs
CMV
R
U5
Ψ
rev
RRE
CMV
Rz
CMV
R
U5
pH6Rz
CMVp
gag
pol
RRE
pA
pGP-RRE3
CMVp
VSV-G
pA
pRV67

Figure 7

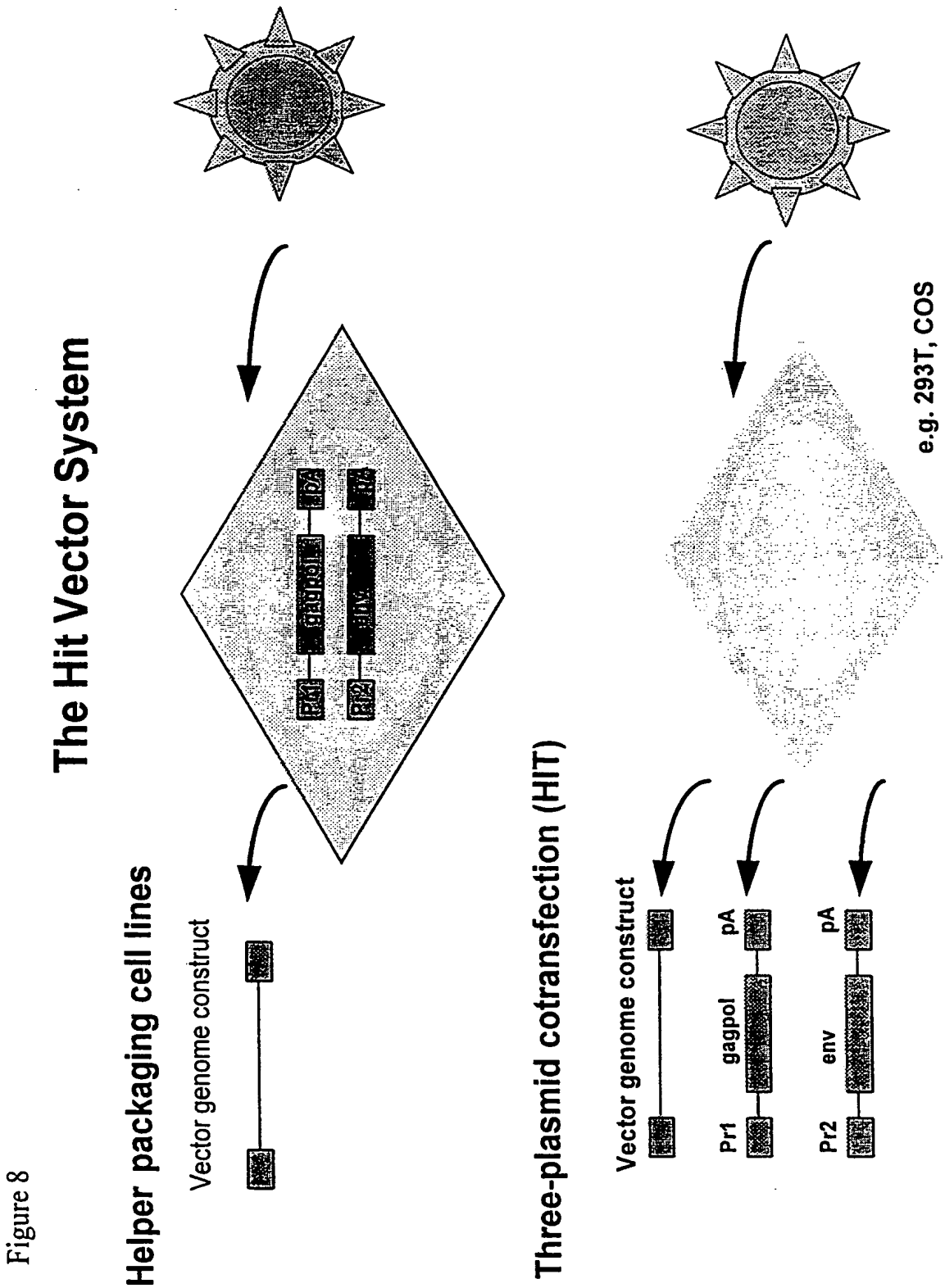
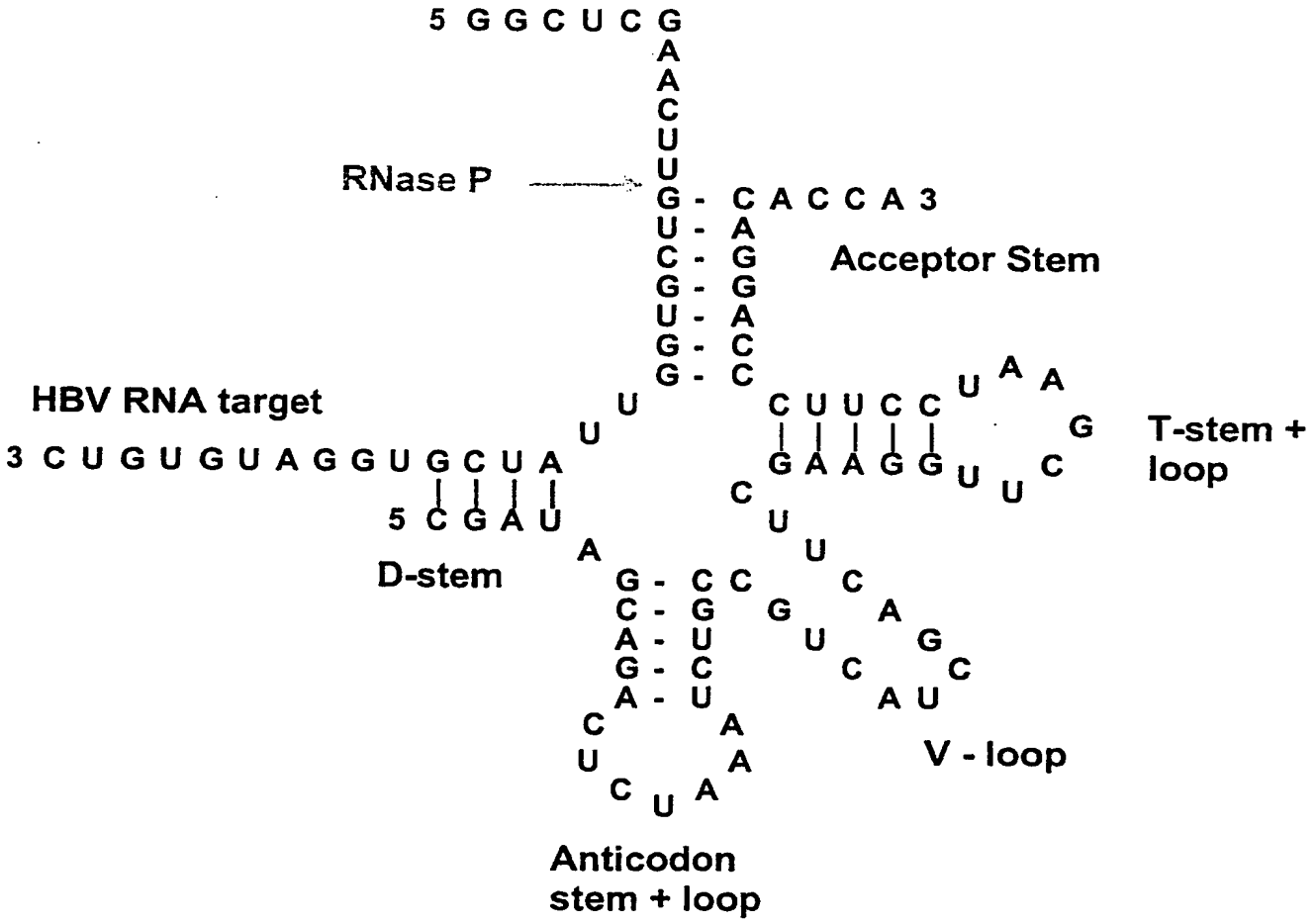

The Hit Vector System
Helper packaging cell lines
Vector genome construct
Three-plasmid cotransfection (HIT)
e.g. 293T, COS
Vector genome construct
pA
gagpol
Pr1
pA
env
Pr2

Figure 8

Figure 9 A



EGS Based on Tyrosyl t-RNA

Figure 9 B

Generic design of EGSs to target any RNA.

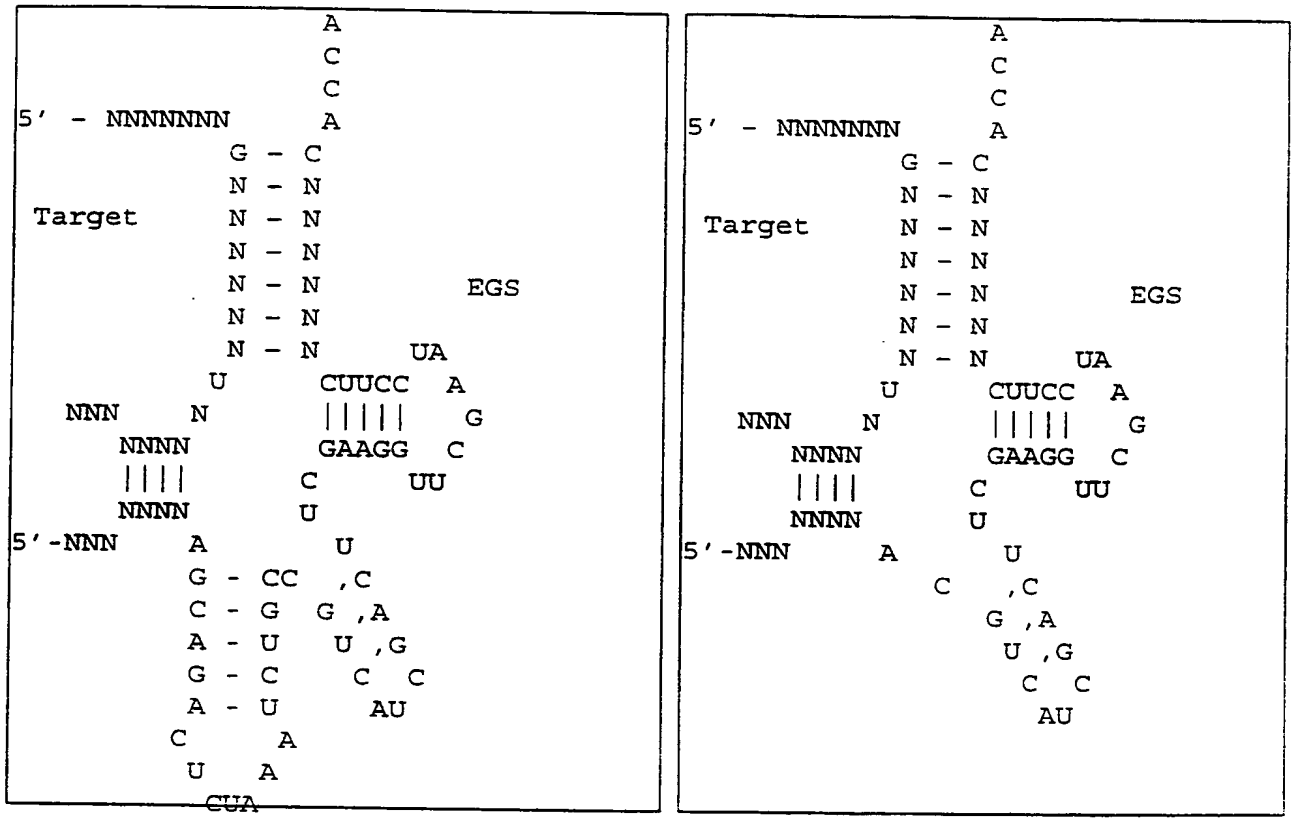
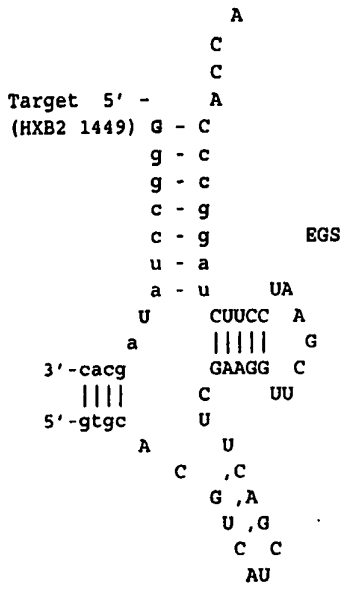
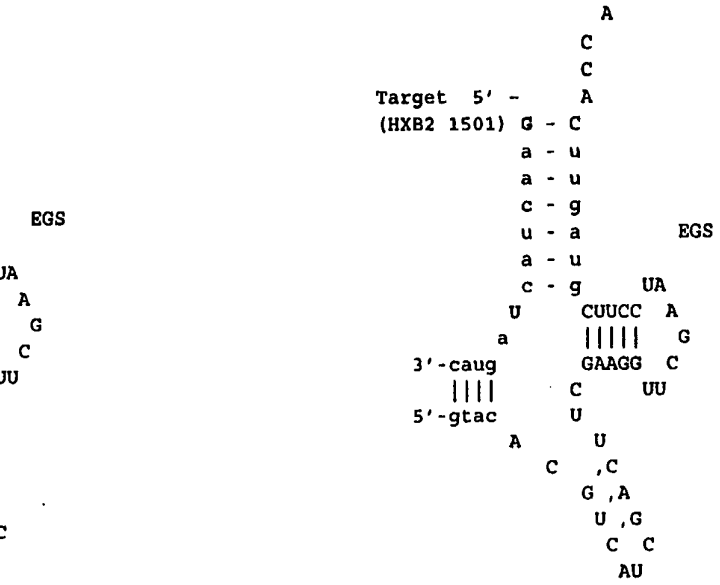
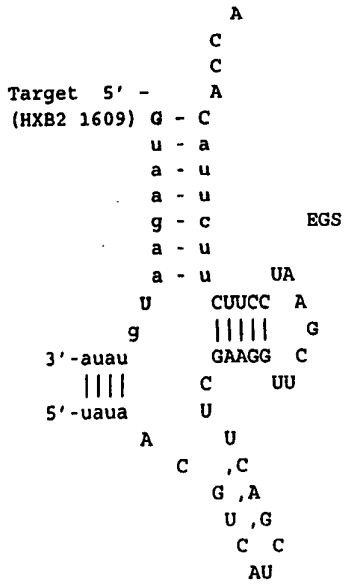


Figure 10 A


Target 5' -
(HXB2 1449)
EGS
3'-cacg
5'-gtgc


Target 5' -
(HXB2 1501)
EGS
3'-caug
5'-gtac


Target 5' -
(HXB2 1609)
EGS
3'-auau
5'-uaua

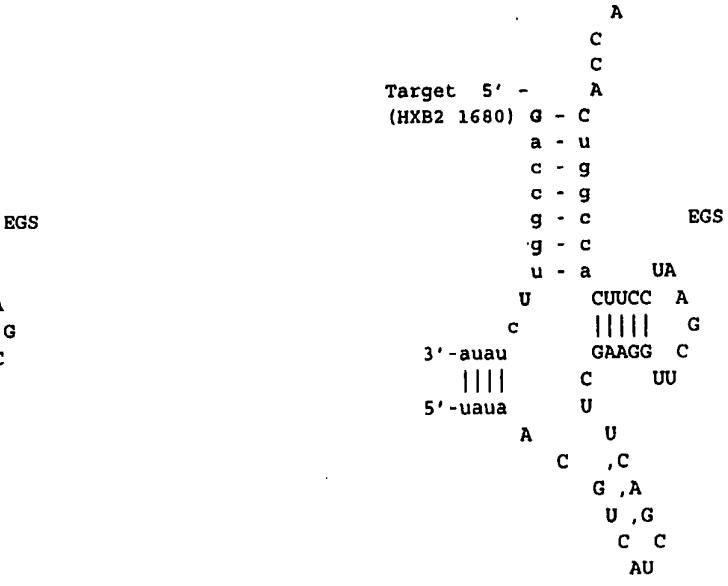
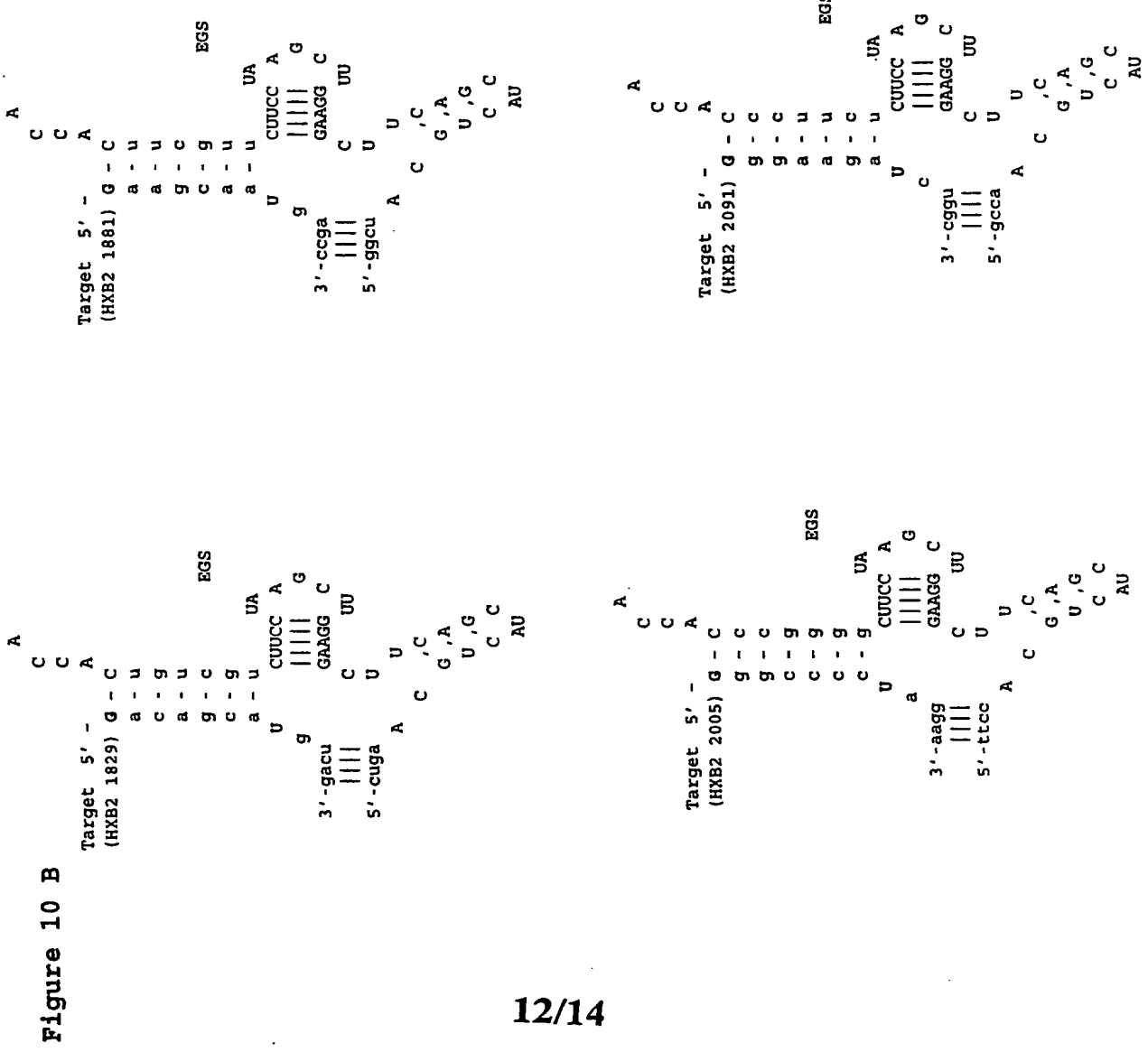
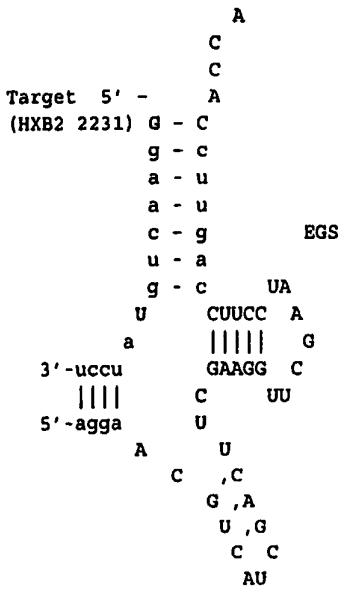
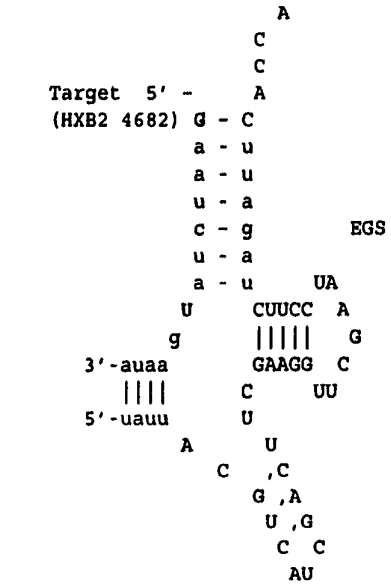

Target 5' -
(HXB2 1680)
EGS
3'-auau
5'-uaua

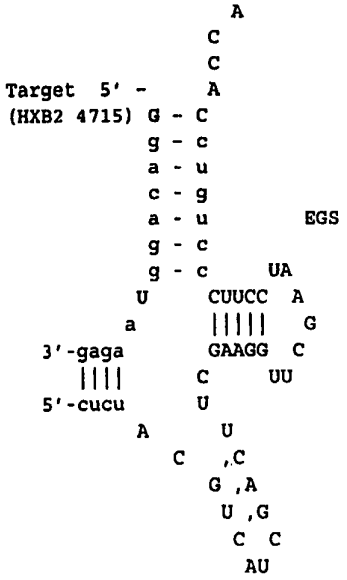
Figure 10 B

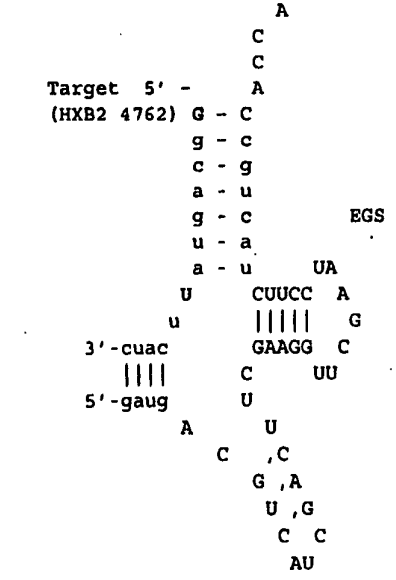

Target 5' - A
(HXB2 1829) G - C
a - u
c - g
a - u
g - c EGS
c - g
a - u UA
U CUUCC A
g ||||| G
3'-gacu GAAGG C
|||| C UU
5'-cuga U
A
C ,C
G ,A
U ,G
C C
AU
Target 5' - A
(HXB2 1881) G - C
a - u
a - u
g - c EGS
c - g
a - u UA
U CUUCC A
g ||||| G
3'-ccga GAAGG C
|||| C UU
5'-ggcu U
A
C ,C
G ,A
U ,G
C C
AU
Target 5' - A
(HXB2 2005) G - C
g - c
g - c
c - g EGS
c - g
c - g UA
U CUUCC A
a ||||| G
3'-aagg GAAGG C
|||| C UU
5'-ttcc U
A
C ,C
G ,A
U ,G
C C
AU
Target 5' - A
(HXB2 2091) G - C
g - c
g - c
a - u EGS
g - c
a - u UA
U CUUCC A
c ||||| G
3'-cggu GAAGG C
|||| C UU
5'-gcca U
A
C ,C
G ,A
U ,G
C C
AU

Figur 10 C


Target 5' -
(HXB2 2231) G - C
g - c
a - u
a - u
c - g
u - a
g - c
EGS
UA
U
CUUCC A
a
G
3'-uccu
GAAGG C
C
UU
5'-agga
U
A
U
C
,C
G ,A
U ,G
C C
AU


Target 5' -
(HXB2 4682) G - C
a - u
a - u
u - a
c - g
u - a
a - u
EGS
UA
U
CUUCC A
g
G
3'-auaa
GAAGG C
C
UU
5'-uauu
U
A
U
C
,C
G ,A
U ,G
C C
AU


Target 5' -
(HXB2 4715) G - C
g - c
a - u
c - g
a - u
g - c
EGS
g - c
UA
U
CUUCC A
a
G
3'-gaga
GAAGG C
C
UU
5'-cucu
U
A
U
C
,C
G ,A
U ,G
C C
AU


Target 5' -
(HXB2 4762) G - C
g - c
c - g
a - u
g - c
EGS
u - a
a - u
UA
U
CUUCC A
u
G
3'-cuac
GAAGG C
C
UU
5'-gaug
U
A
U
C
,C
G ,A
U ,G
C C
AU

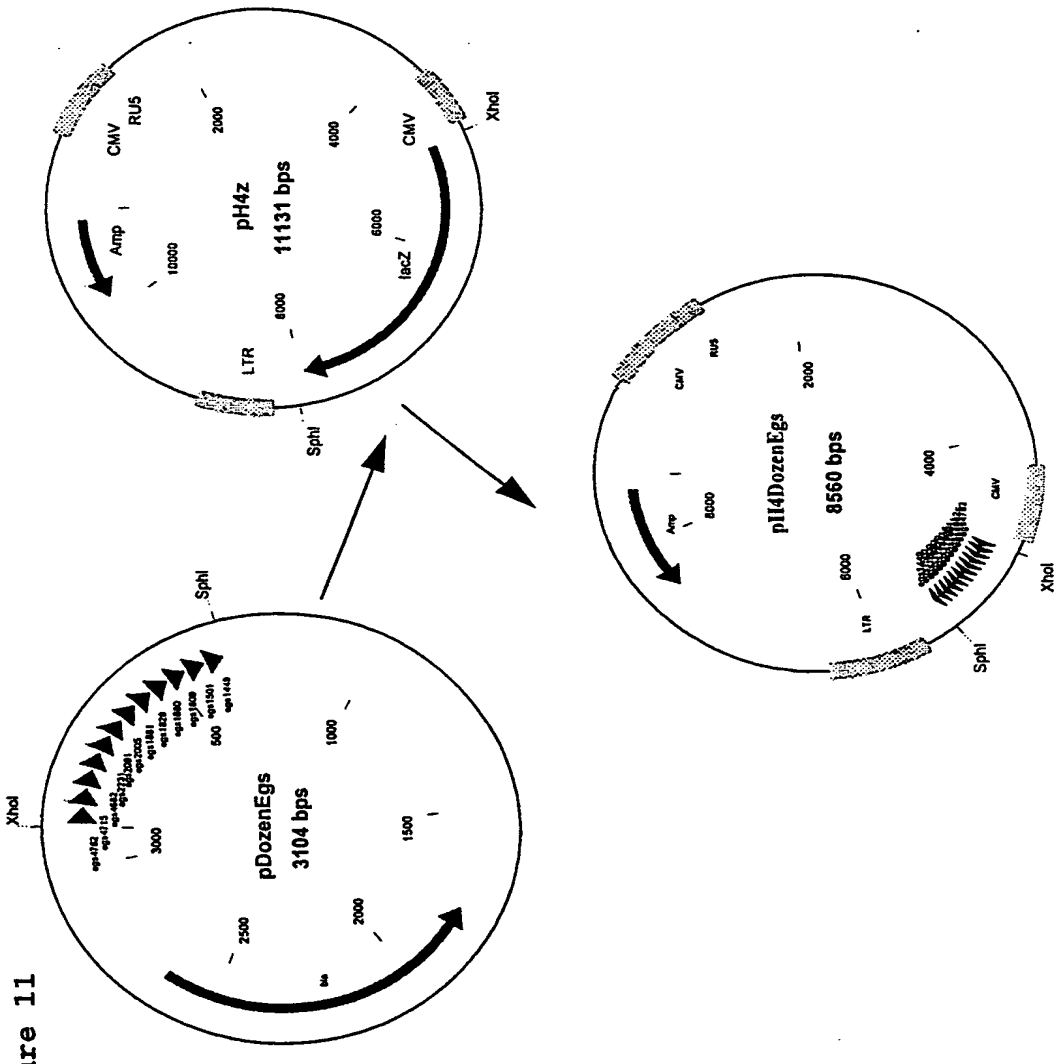

pH4z
11131 bps
CMV
RU5
Amp
LTR
lacZ
CMV
XhoI
SphI
pDozenEgs
3104 bps
XhoI
SphI
pII4DozenEgs
8560 bps
SphI
XhoI

Figure 11

SEQUENCE LISTING PART OF THE DESCRIPTION

SEQ. ID. NO. 1 - Wild type gagpol sequence for strain HXB2 (accession no. K03455)

```
ATGGGTGCGA GAGCGTCAGT ATTAAGCGGG GGAGAATTAG ATCGATGGGA AAAAATTCGG 60
TTAAGGCCAG GGGGAAAGAA AAAATATAAA TTAAAACATA TAGTATGGGC AAGCAGGGAG 120
CTAGAACGAT TCGCAGTTAA TCCTGGCCTG TTAGAAACAT CAGAAGGCTG TAGACAAATA 180
CTGGGACAGC TACAACCATC CCTTCAGACA GGATCAGAAG AACTTAGATC ATTATATAAT 240
ACAGTAGCAA CCCTCTATTG TGTGCATCAA AGGATAGAGA TAAAAGACAC CAAGGAAGCT 300
TTAGACAAGA TAGAGGAAGA GCAAAACAAA AGTAAGAAAA AAGCACAGCA AGCAGCAGCT 360
GACACAGGAC ACAGCAATCA GGTCAGCCAA AATTACCCTA TAGTGCAGAA CATCCAGGGG 420
CAAATGGTAC ATCAGGCCAT ATCACCTAGA ACTTTAAATG CATGGGTAAA AGTAGTAGAA 480
GAGAAGGCTT TCAGCCCAGA AGTGATACCC ATGTTTTCAG CATTATCAGA AGGAGCCACC 540
CCACAAGATT TAAACACCAT GCTAAACACA GTGGGGGGAC ATCAAGCAGC CATGCAAATG 600
TTAAAAGAGA CCATCAATGA GGAAGCTGCA GAATGGGATA GAGTGCATCC AGTGCATGCA 660
GGGCCTATTG CACCAGGCCA GATGAGAGAA CCAAGGGGAA GTGACATAGC AGGAACTACT 720
AGTACCCTTC AGGAACAAAT AGGATGGATG ACAAATAATC CACCTATCCC AGTAGGAGAA 780
ATTTATAAAA GATGGATAAT CCTGGGATTA AATAAAATAG TAAGAATGTA TAGCCCTACC 840
AGCATTCTGG ACATAAGACA AGGACCAAAG GAACCCTTTA GAGACTATGT AGACCGGTTC 900
TATAAAACTC TAAGAGCCGA GCAAGCTTCA CAGGAGGTAA AAAATTGGAT GACAGAAACC 960
TTGTTGGTCC AAAATGCGAA CCCAGATTGT AAGACTATTT TAAAAGCATT GGGACCAGCG 1020
GCTACACTAG AAGAAATGAT GACAGCATGT CAGGGAGTAG GAGGACCCGG CCATAAGGCA 1080
AGAGTTTTGG CTGAAGCAAT GAGCCAAGTA ACAAATTCAG CTACCATAAT GATGCAGAGA 1140
GGCAATTTTA GGAACCAAAG AAAGATTGTT AAGTGTTTCA ATTGTGGCAA AGAAGGGCAC 1200
ACAGCCAGAA ATTGCAGGGC CCCTAGGAAA AAGGGCTGTT GGAAATGTGG AAAGGAAGGA 1260
CACCAAATGA AAGATTGTAC TGAGAGACAG GCTAATTTTT TAGGGAAGAT CTGGCCTTCC 1320
TACAAGGGAA GGCCAGGGAA TTTTCTTCAG AGCAGACCAG AGCCAACAGC CCCACCAGAA 1380
GAGAGCTTCA GGTCTGGGGT AGAGACAACA ACTCCCCCTC AGAAGCAGGA GCCGATAGAC 1440
AAGGAACTGT ATCCTTTAAC TTCCCTCAGG TCACTCTTTG GCAACGACCC CTCGTCACAA 1500
TAAAGATAGG GGGGCAACTA AAGGAAGCTC TATTAGATAC AGGAGCAGAT GATACAGTAT 1560
TAGAAGAAAT GAGTTTGCCA GGAAGATGGA AACCAAAAAT GATAGGGGGA ATTGGAGGTT 1620
TTATCAAAGT AAGACAGTAT GATCAGATAC TCATAGAAAT CTGTGGACAT AAAGCTATAG 1680
GTACAGTATT AGTAGGACCT ACACCTGTCA ACATAATTGG AAGAAATCTG TTGACTCAGA 1740
TTGGTTGCAC TTTAAATTTT CCCATTAGCC CTATTGAGAC TGTACCAGTA AAATTAAAGC 1800
CAGGAATGGA TGGCCCAAAA GTTAAACAAT GGCCATTGAC AGAAGAAAAA ATAAAAGCAT 1860
TAGTAGAAAT TTGTACAGAG ATGGAAAAGG AAGGGAAAAT TTCAAAAATT GGGCCTGAAA 1920
ATCCATACAA TACTCCAGTA TTTGCCATAA AGAAAAAAGA CAGTACTAAA TGGAGAAAAT 1980
TAGTAGATTT CAGAGAACTT AATAAGAGAA CTCAAGACTT CTGGGAAGTT CAATTAGGAA 2040
TACCACATCC CGCAGGGTTA AAAAAGAAAA AATCAGTAAC AGTACTGGAT GTGGGTGATG 2100
CATATTTTTC AGTTCCCTTA GATGAAGACT TCAGGAAGTA TACTGCATTT ACCATACCTA 2160
GTATAAACAA TGAGACACCA GGGATTAGAT ATCAGTACAA TGTGCTTCCA CAGGGATGGA 2220
AAGGATCACC AGCAATATTC CAAAGTAGCA TGACAAAAAT CTTAGAGCCT TTTAGAAAAC 2280
AAAATCCAGA CATAGTTATC TATCAATACA TGGATGATTT GTATGTAGGA TCTGACTTAG 2340
AAATAGGGCA GCATAGAACA AAAATAGAGG AGCTGAGACA ACATCTGTTG AGGTGGGGAC 2400
TTACCACACC AGACAAAAAA CATCAGAAAG AACCTCCATT CCTTTGGATG GGTTATGAAC 2460
TCCATCCTGA TAAATGGACA GTACAGCCTA TAGTGCTGCC AGAAAAAGAC AGCTGGACTG 2520
TCAATGACAT ACAGAAGTTA GTGGGGAAAT TGAATTGGGC AAGTCAGATT TACCCAGGGA 2580
TTAAAGTAAG GCAATTATGT AAACTCCTTA GAGGAACCAA AGCACTAACA GAAGTAATAC 2640
CACTAACAGA AGAAGCAGAG CTAGAACTGG CAGAAAACAG AGAGATTCTA AAAGAACCAG 2700
TACATGGAGT GTATTATGAC CCATCAAAAG ACTTAATAGC AGAAATACAG AAGCAGGGGC 2760
AAGGCCAATG GACATATCAA ATTTATCAAG AGCCATTTAA AAATCTGAAA ACAGGAAAAT 2820
ATGCAAGAAT GAGGGGTGCC CACACTAATG ATGTAAAACA ATTAACAGAG GCAGTGCAAA 2880
AAATAACCAC AGAAAGCATA GTAATATGGG GAAAGACTCC TAAATTTAAA CTGCCCATAC 2940
AAAAGGAAAC ATGGGAAACA TGGTGGACAG AGTATTGGCA AGCCACCTGG ATTCCTGAGT 3000
GGGAGTTTGT TAATACCCCT CCCTTAGTGA AATTATGGTA CCAGTTAGAG AAAGAACCCA 3060
TAGTAGGAGC AGAAACCTTC TATGTAGATG GGGCAGCTAA CAGGGAGACT AAATTAGGAA 3120
AAGCAGGATA TGTTACTAAT AGAGGAAGAC AAAAAGTTGT CACCCTAACT GACACAACAA 3180
ATCAGAAGAC TGAGTTACAA GCAATTTATC TAGCTTTGCA GGATTCGGGA TTAGAAGTAA 3240
ACATAGTAAC AGACTCACAA TATGCATTAG GAATCATTCA AGCACAACCA GATCAAAGTG 3300
AATCAGAGTT AGTCAATCAA ATAATAGAGC AGTTAATAAA AAAGGAAAAG GTCTATCTGG 3360
CATGGGTACC AGCACACAAA GGAATTGGAG GAAATGAACA AGTAGATAAA TTAGTCAGTG 3420
CTGGAATCAG GAAAGTACTA TTTTTAGATG GAATAGATAA GGCCCAAGAT GAACATGAGA 3480
AATATCACAG TAATTGGAGA GCAATGGCTA GTGATTTTAA CCTGCCACCT GTAGTAGCAA 3540
```

AAGAAATAGT AGCCAGCTGT GATAAATGTC AGCTAAAAGG AGAAGCCATG CATGGACAAG 3600
TAGACTGTAG TCCAGGAATA TGGCAACTAG ATTGTACACA TTTAGAAGGA AAAGTTATCC 3660
TGGTAGCAGT TCATGTAGCC AGTGGATATA TAGAAGCAGA AGTTATTCCA GCAGAAACAG 3720
GGCAGGAAAC AGCATATTTT CTTTTAAAAT TAGCAGGAAG ATGGCCAGTA AAAACAATAC 3780
ATACTGACAA TGGCAGCAAT TTCACCGGTG CTACGGTTAG GGCCGCCTGT TGGTGGGCGG 3840
GAATCAAGCA GGAATTTGGA ATTCCCTACA ATCCCCAAAG TCAAGGAGTA GTAGAATCTA 3900
TGAATAAAGA ATTAAAGAAA ATTATAGGAC AGGTAAGAGA TCAGGCTGAA CATCTTAAGA 3960
CAGCAGTACA AATGGCAGTA TTCATCCACA ATTTTAAAAG AAAAGGGGGG ATTGGGGGGT 4020
ACAGTGCAGG GGAAAGAATA GTAGACATAA TAGCAACAGA CATACAAACT AAAGAATTAC 4080
AAAAACAAAT TACAAAAATT CAAAATTTTC GGGTTTATTA CAGGGACAGC AGAAATTCAC 4140
TTTGGAAAGG ACCAGCAAAG CTCCTCTGGA AAGGTGAAGG GGCAGTAGTA ATACAAGATA 4200
ATAGTGACAT AAAAGTAGTG CCAAGAAGAA AAGCAAAGAT CATTAGGGAT TATGGAAAAC 4260
AGATGGCAGG TGATGATTGT GTGGCAAGTA GACAGGATGA GGATTAG 4307

SEQ I.D. NO. 2 - gagpol-SYNgp - codon optimised gagpol sequence

ATGGGCGCCC GCGCCAGCGT GCTGTCGGGC GGCGAGCTGG ACCGCTGGGA GAAGATCCGC 60
CTGCGCCCCG GCGGCAAAAA GAAGTACAAG CTGAAGCACA TCGTGTGGGC CAGCCGCGAA 120
CTGGAGCGCT TCGCCGTGAA CCCCGGGCTC CTGGAGACCA GCGAGGGGTG CCGCCAGATC 180
CTCGGCCAAC TGCAGCCCAG CCTGCAAACC GGCAGCGAGG AGCTGCGCAG CCTGTACAAC 240
ACCGTGGCCA CGCTGTACTG CGTCCACCAG CGCATCGAAA TCAAGGATAC GAAAGAGGCC 300
CTGGATAAAA TCGAAGAGGA ACAGAATAAG AGCAAAAAGA AGGCCCAACA GGCCGCCGCG 360
GACACCGGAC ACAGCAACCA GGTCAGCCAG AACTACCCCA TCGTGCAGAA CATCCAGGGG 420
CAGATGGTGC ACCAGGCCAT CTCCCCCCGC ACGCTGAACG CCTGGGTGAA GGTGGTGGAA 480
GAGAAGGCTT TTAGCCCGGA GGTGATACCC ATGTTCTCAG CCCTGTCAGA GGGAGCCACC 540
CCCCAAGATC TGAACACCAT GCTCAACACA GTGGGGGGAC ACCAGGCCGC CATGCAGATG 600
CTGAAGGAGA CCATCAATGA GGAGGCTGCC GAATGGGATC GTGTGCATCC GGTGCACGCA 660
GGGCCCATCG CACCGGGCCA GATGCGTGAG CCACGGGGCT CAGACATCGC CGGAACGACT 720
AGTACCCTTC AGGAACAGAT CGGCTGGATG ACCAACAACC CACCCATCCC GGTGGGAGAA 780
ATCTACAAAC GCTGGATCAT CCTGGGCCTG AACAAGATCG TGCGCATGTA TAGCCCTACC 840
AGCATCCTGG ACATCCGCCA AGGCCCGAAG GAACCCTTTC GCGACTACGT GGACCGGTTC 900
TACAAAACGC TCCGCGCCGA GCAGGCTAGC CAGGAGGTGA AGAACTGGAT GACCGAAACC 960
CTGCTGGTCC AGAACGCGAA CCCGGACTGC AAGACGATCC TGAAGGCCCT GGGCCCAGCG 1020
GCTACCCTAG AGGAAATGAT GACCGCCTGT CAGGGAGTGG GCGGACCCGG CCACAAGGCA 1080
CGCGTCCTGG CTGAGGCCAT GAGCCAGGTG ACCAACTCCG CTACCATCAT GATGCAGCGC 1140
GGCAACTTTC GGAACCAACG CAAGATCGTC AAGTGCTTCA ACTGTGGCAA AGAAGGGCAC 1200
ACAGCCCGCA ACTGCAGGGC CCCTAGGAAA AAGGGCTGCT GGAAATGCGG CAAGGAAGGC 1260
CACCAGATGA AAGACTGTAC TGAGAGACAG GCTAATTTTT TAGGGAAGAT CTGGCCTTCC 1320
TACAAGGGAA GGCCAGGGAA TTTTCTTCAG AGCAGACCAG AGCCAACAGC CCCACCAGAA 1380
GAGAGCTTCA GGTCTGGGGT AGAGACAACA ACTCCCCCTC AGAAGCAGGA GCCGATAGAC 1440
AAGGAACTGT ATCCTTTAAC TTCCCTCAGA TCACTCTTTG GCAACGACCC CTCGTCACAA 1500
TAAAGATAGG GGGGCAGCTC AAGGAGGCTC TCCTGGACAC CGGAGCAGAC GACACCGTGC 1560
TGGAGGAGAT GTCGTTGCCA GGCCGCTGGA AGCCGAAGAT GATCGGGGGA ATCGGCGGTT 1620
TCATCAAGGT GCGCCAGTAT GACCAGATCC TCATCGAAAT CTGCGGCCAC AAGGCTATCG 1680
GTACCGTGCT GGTGGGCCCC ACACCCGTCA ACATCATCGG ACGCAACCTG TTGACGCAGA 1740
TCGGTTGCAC GCTGAACTTC CCCATTAGCC CTATCGAGAC GGTACCGGTG AAGCTGAAGC 1800
CCGGGATGGA CGGCCCGAAG GTCAAGCAAT GGCCATTGAC AGAGGAGAAG ATCAAGGCAC 1860
TGGTGGAGAT TTGCACAGAG ATGGAAAAGG AAGGGAAAAT CTCCAAGATT GGGCCTGAGA 1920
ACCCGTACAA CACGCCGGTG TTCGCAATCA AGAAGAAGGA CTCGACGAAA TGGCGCAAGC 1980
TGGTGGACTT CCGCGAGCTG AACAAGCGCA CGCAAGACTT CTGGGAGGTT CAGCTGGGCA 2040
TCCCGCACCC CGCAGGGCTG AAGAAGAAGA AATCCGTGAC CGTACTGGAT GTGGGTGATG 2100
CCTACTTCTC CGTTCCCCTG GACGAAGACT TCAGGAAGTA CACTGCCTTC ACAATCCCTT 2160
CGATCAACAA CGAGACACCG GGGATTCGAT ATCAGTACAA CGTGCTGCCC CAGGGCTGGA 2220
AAGGCTCTCC CGCAATCTTC CAGAGTAGCA TGACCAAAAT CCTGGAGCCT TTCCGCAAAC 2280
AGAACCCCGA CATCGTCATC TATCAGTACA TGGATGACTT GTACGTGGGC TCTGATCTAG 2340
AGATAGGGCA GCACCGCACC AAGATCGAGG AGCTGCGCCA GCACCTGTTG AGGTGGGGAC 2400
TGACCACACC CGACAAGAAG CACCAGAAGG AGCCTCCCTT CCTCTGGATG GGTTACGAGC 2460
TGCACCCTGA CAAATGGACC GTGCAGCCTA TCGTGCTGCC AGAGAAAGAC AGCTGGACTG 2520
TCAACGACAT ACAGAAGCTG GTGGGGAAGT TGAACTGGGC CAGTCAGATT TACCCAGGGA 2580
TTAAGGTGAG GCAGCTGTGC AAACTCCTCC GCGGAACCAA GGCACTCACA GAGGTGATCC 2640
CCCTAACCGA GGAGGCCGAG CTCGAACTGG CAGAAAACCG AGAGATCCTA AAGGAGCCCG 2700
TGCACGGCGT GTACTATGAC CCCTCCAAGG ACCTGATCGC CGAGATCCAG AAGCAGGGGC 2760
AAGGCCAGTG GACCTATCAG ATTTACCAGG AGCCCTTCAA GAACCTGAAG ACCGGCAAGT 2820
ACGCCCGGAT GAGGGGTGCC CACACTAACG ACGTCAAGCA GCTGACCGAG GCCGTGCAGA 2880

```
AGATCACCAC CGAAAGCATC GTGATCTGGG GAAAGACTCC TAAGTTCAAG CTGCCCATCC 2940
AGAAGGAAAC CTGGGAAACC TGGTGGACAG AGTATTGGCA GGCCACCTGG ATTCCTGAGT 3000
GGGAGTTCGT CAACACCCCT CCCCTGGTGA AGCTGTGGTA CCAGCTGGAG AAGGAGCCCA 3060
TAGTGGGCGC CGAAACCTTC TACGTGGATG GGGCCGCTAA CAGGGAGACT AAGCTGGGCA 3120
AAGCCGGATA CGTCACTAAC CGGGGCAGAC AGAAGGTTGT CACCCTCACT GACACCACCA 3180
ACCAGAAGAC TGAGCTGCAG GCCATTTACC TCGCTTTGCA GGACTCGGGC CTGGAGGTGA 3240
ACATCGTGAC AGACTCTCAG TATGCCCTGG GCATCATTCA AGCCCAGCCA GACCAGAGTG 3300
AGTCCGAGCT GGTCAATCAG ATCATCGAGC AGCTGATCAA GAAGGAAAAG GTCTATCTGG 3360
CCTGGGTACC CGCCCACAAA GGCATTGGCG GCAATGAGCA GGTCGACAAG CTGGTCTCGG 3420
CTGGCATCAG GAAGGTGCTA TTCCTGGATG GCATCGACAA GGCCCAGGAC GAGCACGAGA 3480
AATACCACAG CAACTGGCGG GCCATGGCTA GCGACTTCAA CCTGCCCCCT GTGGTGGCCA 3540
AAGAGATCGT GGCCAGCTGT GACAAGTGTC AGCTCAAGGG CGAAGCCATG CATGGCCAGG 3600
TGGACTGTAG CCCCGGCATC TGGCAACTCG ATTGCACCCA TCTGGAGGGC AAGGTTATCC 3660
TGGTAGCCGT CCATGTGGCC AGTGGCTACA TCGAGGCCGA GGTCATTCCC GCCGAAACAG 3720
GGCAGGAGAC AGCCTACTTC CTCCTGAAGC TGGCAGGCCG GTGGCCAGTG AAGACCATCC 3780
ATACTGACAA TGGCAGCAAT TTCACCAGTG CTACGGTTAA GGCCGCCTGC TGGTGGGCGG 3840
GAATCAAGCA GGAGTTCGGG ATCCCCTACA ATCCCCAGAG TCAGGGCGTC GTCGAGTCTA 3900
TGAATAAGGA GTTAAAGAAG ATTATCGGCC AGGTCAGAGA TCAGGCTGAG CATCTCAAGA 3960
CCGCGGTCCA AATGGCGGTA TTCATCCACA ATTTCAAGCG GAAGGGGGGG ATTGGGGGGT 4020
ACAGTGCGGG GGAGCGGATC GTGGACATCA TCGCGACCGA CATCCAGACT AAGGAGCTGC 4080
AAAAGCAGAT TACCAAGATT CAGAATTTCC GGGTCTACTA CAGGGACAGC AGAAATCCCC 4140
TCTGGAAAGG CCCAGCGAAG CTCCTCTGGA AGGGTGAGGG GGCAGTAGTG ATCCAGGATA 4200
ATAGCGACAT CAAGGTGGTG CCCAGAAGAA AGGCGAAGAT CATTAGGGAT TATGGCAAAC 4260
AGATGGCGGG TGATGATTGC GTGGCGAGCA GACAGGATGA GGATTAG                 4307
```

SEQ. ID. NO. 3 - Envelope Gene from HIV-1 MN (Genbank accession no. M17449)

```
ATGAGAGTGA AGGGGATCAG GAGGAATTAT CAGCACTGGT GGGGATGGGG CACGATGCTC 60
CTTGGGTTAT TAATGATCTG TAGTGCTACA GAAAAATTGT GGGTCACAGT CTATTATGGG 120
GTACCTGTGT GGAAAGAAGC AACCACCACT CTATTTTGTG CATCAGATGC TAAAGCATAT 180
GATACAGAGG TACATAATGT TTGGGCCACA CAAGCCTGTG TACCCACAGA CCCCAACCCA 240
CAAGAAGTAG AATTGGTAAA TGTGACAGAA AATTTTAACA TGTGGAAAAA TAACATGGTA 300
GAACAGATGC ATGAGGATAT AATCAGTTTA TGGGATCAAA GCCTAAAGCC ATGTGTAAAA 360
TTAACCCCAC TCTGTGTTAC TTTAAATTGC ACTGATTTGA GGAATACTAC TAATACCAAT 420
AATAGTACTG CTAATAACAA TAGTAATAGC GAGGGAACAA TAAAGGGAGG AGAAATGAAA 480
AACTGCTCTT TCAATATCAC CACAAGCATA AGAGATAAGA TGCAGAAAGA ATATGCACTT 540
CTTTATAAAC TTGATATAGT ATCAATAGAT AATGATAGTA CCAGCTATAG GTTGATAAGT 600
TGTAATACCT CAGTCATTAC ACAAGCTTGT CCAAAGATAT CCTTTGAGCC AATTCCCATA 660
CACTATTGTG CCCCGGCTGG TTTTGCGATT CTAAAATGTA ACGATAAAAA GTTCAGTGGA 720
AAAGGATCAT GTAAAAATGT CAGCACAGTA CAATGTACAC ATGGAATTAG GCCAGTAGTA 780
TCAACTCAAC TGCTGTTAAA TGGCAGTCTA GCAGAAGAAG AGGTAGTAAT TAGATCTGAG 840
AATTTCACTG ATAATGCTAA AACCATCATA GTACATCTGA ATGAATCTGT ACAAATTAAT 900
TGTACAAGAC CCAACTACAA TAAAAGAAAA AGGATACATA TAGGACCAGG GAGAGCATTT 960
TATACAACAA AAAATATAAT AGGAACTATA AGACAAGCAC ATTGTAACAT TAGTAGAGCA 1020
AAATGGAATG ACACTTTAAG ACAGATAGTT AGCAAATTAA AAGAACAATT TAAGAATAAA 1080
ACAATAGTCT TTAATCAATC CTCAGGAGGG GACCCAGAAA TTGTAATGCA CAGTTTTAAT 1140
TGTGGAGGGG AATTTTTCTA CTGTAATACA TCACCACTGT TTAATAGTAC TTGGAATGGT 1200
AATAATACTT GGAATAATAC TACAGGGTCA AATAACAATA TCACACTTCA ATGCAAAATA 1260
AAACAAATTA TAAACATGTG GCAGGAAGTA GGAAAAGCAA TGTATGCCCC TCCCATTGAA 1320
GGACAAATTA GATGTTCATC AAATATTACA GGGCTACTAT TAACAAGAGA TGGTGGTAAG 1380
GACACGGACA CGAACGACAC CGAGATCTTC AGACCTGGAG GAGGAGATAT GAGGGACAAT 1440
TGGAGAAGTG AATTATATAA ATATAAAGTA GTAACAATTG AACCATTAGG AGTAGCACCC 1500
ACCAAGGCAA AGAGAAGAGT GGTGCAGAGA GAAAAAAGAG CAGCGATAGG AGCTCTGTTC 1560
CTTGGGTTCT TAGGAGCAGC AGGAAGCACT ATGGGCGCAG CGTCAGTGAC GCTGACGGTA 1620
CAGGCCAGAC TATTATTGTC TGGTATAGTG CAACAGCAGA ACAATTTGCT GAGGGCCATT 1680
GAGGCGCAAC AGCATATGTT GCAACTCACA GTCTGGGGCA TCAAGCAGCT CCAGGCAAGA 1740
GTCCTGGCTG TGGAAAGATA CCTAAAGGAT CAACAGCTCC TGGGGTTTTG GGGTTGCTCT 1800
GGAAAACTCA TTTGCACCAC TACTGTGCCT TGGAATGCTA GTTGGAGTAA TAAATCTCTG 1860
GATGATATTT GGAATAACAT GACCTGGATG CAGTGGGAAA GAGAAATTGA CAATTACACA 1920
AGCTTAATAT ACTCATTACT AGAAAAATCG CAAACCCAAC AAGAAAAGAA TGAACAAGAA 1980
TTATTGGAAT TGGATAAATG GGCAAGTTTG TGGAATTGGT TTGACATAAC AAATTGGCTG 2040
TGGTATATAA AAATATTCAT AATGATAGTA GGAGGCTTGG TAGGTTTAAG AATAGTTTTT 2100
GCTGTACTTT CTATAGTGAA TAGAGTTAGG CAGGGATACT CACCATTGTC GTTGCAGACC 2160
CGCCCCCCAG TTCCGAGGGG ACCCGACAGG CCCGAAGGAA TCGAAGAAGA AGGTGGAGAG 2220
```

```
AGAGACAGAG ACACATCCGG TCGATTAGTG CATGGATTCT TAGCAATTAT CTGGGTCGAC 2280
CTGCGGAGCC TGTTCCTCTT CAGCTACCAC CACAGAGACT TACTCTTGAT TGCAGCGAGG 2340
ATTGTGGAAC TTCTGGGACG CAGGGGGTGG GAAGTCCTCA AATATTGGTG GAATCTCCTA 2400
CAGTATTGGA GTCAGGAACT AAAGAGTAGT GCTGTTAGCT TGCTTAATGC CACAGCTATA 2460
GCAGTAGCTG AGGGGACAGA TAGGGTTATA GAAGTACTGC AAAGAGCTGG TAGAGCTATT 2520
CTCCACATAC CTACAAGAAT AAGACAGGGC TTGGAAAGGG CTTTGCTATA A           2571
```

SEQ. I.D. NO. 4 - SYNgp-160mn - codon optimised env sequence

```
ATGAGGGTGA AGGGGATCCG CCGCAACTAC CAGCACTGGT GGGGCTGGGG CACGATGCTC 60
CTGGGGCTGC TGATGATCTG CAGCGCCACC GAGAAGCTGT GGGTGACCGT GTACTACGGC 120
GTGCCCGTGT GGAAGGAGGC CACCACCACC CTGTTCTGCG CCAGCGACGC CAAGGCGTAC 180
GACACCGAGG TGCACAACGT GTGGGCCACC CAGGCGTGCG TGCCCACCGA CCCCAACCCC 240
CAGGAGGTGG AGCTCGTGAA CGTGACCGAG AACTTCAACA TGTGGAAGAA CAACATGGTG 300
GAGCAGATGC ATGAGGACAT CATCAGCCTG TGGGACCAGA GCCTGAAGCC CTGCGTGAAG 360
CTGACCCCCC TGTGCGTGAC CCTGAACTGC ACCGACCTGA GGAACACCAC CAACACCAAC 420
AACAGCACCG CCAACAACAA CAGCAACAGC GAGGGCACCA TCAAGGGCGG CGAGATGAAG 480
AACTGCAGCT TCAACATCAC CACCAGCATC CGCGACAAGA TGCAGAAGGA GTACGCCCTG 540
CTGTACAAGC TGGATATCGT GAGCATCGAC AACGACAGCA CCAGCTACCG CCTGATCTCC 600
TGCAACACCA GCGTGATCAC CCAGGCCTGC CCCAAGATCA GCTTCGAGCC CATCCCCATC 660
CACTACTGCG CCCCCGCCGG CTTCGCCATC CTGAAGTGCA ACGACAAGAA GTTCAGCGGC 720
AAGGGCAGCT GCAAGAACGT GAGCACCGTG CAGTGCACCC ACGGCATCCG GCCGGTGGTG 780
AGCACCCAGC TCCTGCTGAA CGGCAGCCTG GCCGAGGAGG AGGTGGTGAT CCGCAGCGAG 840
AACTTCACCG ACAACGCCAA GACCATCATC GTGCACCTGA ATGAGAGCGT GCAGATCAAC 900
TGCACGCGTC CCAACTACAA CAAGCGCAAG CGCATCCACA TCGGCCCCGG GCGCGCCTTC 960
TACACCACCA AGAACATCAT CGGCACCATC CGCCAGGCCC ACTGCAACAT CTCTAGAGCC 1020
AAGTGGAACG ACACCCTGCG CCAGATCGTG AGCAAGCTGA AGGAGCAGTT CAAGAACAAG 1080
ACCATCGTGT TCAACCAGAG CAGCGGCGGC GACCCCGAGA TCGTGATGCA CAGCTTCAAC 1140
TGCGGCGGCG AATTCTTCTA CTGCAACACC AGCCCCCTGT TCAACAGCAC CTGGAACGGC 1200
AACAACACCT GGAACAACAC CACCGGCAGC AACAACAATA TTACCCTCCA GTGCAAGATC 1260
AAGCAGATCA TCAACATGTG GCAGGAGGTG GGCAAGGCCA TGTACGCCCC CCCCATCGAG 1320
GGCCAGATCC GGTGCAGCAG CAACATCACC GGTCTGCTGC TGACCCGCGA CGGCGGCAAG 1380
GACACCGACA CCAACGACAC CGAAATCTTC CGCCCCGGCG GCGGCGACAT GCGCGACAAC 1440
TGGAGATCTG AGCTGTACAA GTACAAGGTG GTGACGATCG AGCCCCTGGG CGTGGCCCCC 1500
ACCAAGGCCA AGCGCCGCGT GGTGCAGCGC GAGAAGCGGG CCGCCATCGG CGCCCTGTTC 1560
CTGGGCTTCC TGGGGGCGGC GGGCAGCACC ATGGGGGCCG CCAGCGTGAC CCTGACCGTG 1620
CAGGCCCGCC TGCTCCTGAG CGGCATCGTG CAGCAGCAGA ACAACCTCCT CCGCGCCATC 1680
GAGGCCCAGC AGCATATGCT CCAGCTCACC GTGTGGGGCA TCAAGCAGCT CCAGGCCCGC 1740
GTGCTGGCCG TGGAGCGCTA CCTGAAGGAC CAGCAGCTCC TGGGCTTCTG GGGCTGCTCC 1800
GGCAAGCTGA TCTGCACCAC CACGGTACCC TGGAACGCCT CCTGGAGCAA CAAGAGCCTG 1860
GACGACATCT GGAACAACAT GACCTGGATG CAGTGGGAGC GCGAGATCGA TAACTACACC 1920
AGCCTGATCT ACAGCCTGCT GGAGAAGAGC CAGACCCAGC AGGAGAAGAA CGAGCAGGAG 1980
CTGCTGGAGC TGGACAAGTG GGCGAGCCTG TGGAACTGGT TCGACATCAC CAACTGGCTG 2040
TGGTACATCA AAATCTTCAT CATGATTGTG GGCGGCCTGG TGGGCCTCCG CATCGTGTTC 2100
GCCGTGCTGA GCATCGTGAA CCGCGTGCGC CAGGGCTACA GCCCCCTGAG CCTCCAGACC 2160
CGGCCCCCCG TGCCGCGCGG GCCCGACCGC CCCGAGGGCA TCGAGGAGGA GGGCGGCGAG 2220
CGCGACCGCG ACACCAGCGG CAGGCTCGTG CACGGCTTCC TGGCGATCAT CTGGGTCGAC 2280
CTCCGCAGCC TGTTCCTGTT CAGCTACCAC CACCGCGACC TGCTGCTGAT CGCCGCCCGC 2340
ATCGTGGAAC TCCTAGGCCG CCGCGGCTGG GAGGTGCTGA AGTACTGGTG GAACCTCCTC 2400
CAGTATTGGA GCCAGGAGCT GAAGTCCAGC GCCGTGAGCC TGCTGAACGC CACCGCCATC 2460
GCCGTGGCCG AGGGCACCGA CCGCGTGATC GAGGTGCTCC AGAGGGCCGG GAGGGCGATC 2520
CTGCACATCC CCACCCGCAT CCGCCAGGGG CTCGAGAGGG CGCTGCTGTA A           2571
```

SEQ. I.D. NO. 11 - Complete Sequence of pH4DOZENEGS

```
CTGACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG CGCAGCGTGA 60
CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTCGC TTTCTTCCCT TCCTTTCTCG 120
CCACGTTCGC CGGCTTTCCC CGTCAAGCTC TAAATCGGGG GCTCCCTTTA GGGTTCCGAT 180
TTAGTGCTTT ACGGCACCTC GACCCCAAAA AACTTGATTA GGGTGATGGT TCACGTAGTG 240
GGCCATCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTCTTTAATA 300
GTGGACTCTT GTTCCAAACT GGAACAACAC TCAACCCTAT CTCGGTCTAT TCTTTTGATT 360
TATAAGGGAT TTTGCCGATT TCGGCCTATT GGTTAAAAAA TGAGCTGATT TAACAAAAAT 420
TTAACGCGAA TTTTAACAAA ATATTAACGC TTACAATTTC CATTCGCCAT TCAGGCTGCG 480
```

CAACTGTTGG GAAGGGCGAT CGGTGCGGGC CTCTTCGCTA TTACGCCAGC TGGCGAAAGG 540
GGGATGTGCT GCAAGGCGAT TAAGTTGGGT AACGCCAGGG TTTTCCCAGT CACGACGTTG 600
TAAAACGACG GCCAGTGAGC GCGCGTAATA CGACTCACTA TAGGGCGAAT TGGAGCTCCA 660
CCGCGGTGGC GGCCGCTCTA GAGTCCGTTA CATAACTTAC GGTAAATGGC CCGCCTGGCT 720
GACCGCCCAA CGACCCCCGC CCATTGACGT CAATAATGAC GTATGTTCCC ATAGTAACGC 780
CAATAGGGAC TTTCCATTGA CGTCAATGGG TGGAGTATTT ACGGTAAACT GCCCACTTGG 840
CAGTACATCA AGTGTATCAT ATGCCAAGTA CGCCCCCTAT TGACGTCAAT GACGGTAAAT 900
GGCCCGCCTG GCATTATGCC CAGTACATGA CCTTATGGGA CTTTCCTACT TGGCAGTACA 960
TCTACGTATT AGTCATCGCT ATTACCATGG TGATGCGGTT TTGGCAGTAC ATCAATGGGC 1020
GTGGATAGCG GTTTGACTCA CGGGGATTTC CAAGTCTCCA CCCCATTGAC GTCAATGGGA 1080
GTTTGTTTTG GCACCAAAAT CAACGGGACT TTCCAAAATG TCGTAACAAC TCCGCCCCAT 1140
TGACGCAAAT GGGCGGTAGG CGTGTACGGT GGGAGGTCTA TATAAGCAGA GCTCGTTTAG 1200
TGAACCGGTC TCTCTGGTTA GACCAGATCT GAGCCTGGGA GCTCTCTGGC TAACTAGGGA 1260
ACCCACTGCT TAAGCCTCAA TAAAGCTTGC CTTGAGTGCT TCAAGTAGTG TGTGCCCGTC 1320
TGTTGTGTGA CTCTGGTAAC TAGAGATCCC TCAGACCCTT TTAGTCAGTG TGGAAAATCT 1380
CTAGCAGTGG CGCCCGAACA GGGACTTGAA AGCGAAAGGG AAACCAGAGG AGCTCTCTCG 1440
ACGCAGGACT CGGCTTGCTG AAGCGCGCAC GGCAAGAGGC GAGGGGCGGC GACTGGTGAG 1500
TACGCCAAAA ATTTTGACTA GCGGAGGCTA GAAGGAGAGA GATGGGTGCG AGAGCGTCAG 1560
TATTAAGCGG GGGAGAATTA GATCGCGATG GGAAAAAATT CGGTTAAGGC CAGGGGGAAA 1620
GAAAAAATAT AAATTAAAAC ATATAGTATG GGCAAGCAGG GAGCTAGAAC GATTCGCAGT 1680
TAATCCTGGC CTGTTAGAAA CATCAGAAGG CTGTAGACAA ATACTGGGAC AGCTACAACC 1740
ATCCCTTCAG ACAGGATCAG AAGAACTTAG ATCATTATAT AATACAGTAG CAACCCTCTA 1800
TTGTGTGCAT CAAAGGTTGA GATAAAAGAC ACCAAGGAAG CTTTAGACAA GATAGAGGGA 1860
GAGCAAAACA AAAGTAAGAA AAAAGCACAG CAAGCAGCAG CTGACACAGG ACACAGCAAT 1920
CAGGTCAGCC AAAATTACCC TATAGTGCAG AACATCCAGG GGCAAATGGT ACATCAGGCC 1980
ATATCACCTA GAACTTTAAA TGCATGGGTA AAAGTAGTAG AAGAGAAGGC TTTCAGCCCA 2040
GAAGTGATAC CCATGTTTTC AGCATTATCA GAAGGAGCCA CCCCACAAGA TTTAAACACC 2100
ATGCTAAACA CAGTGGGGGG ACATCAAGCA GCCATGCAAA TGTTAAAAGA GACCATCAAT 2160
GAGGAAGCTG CAGGAATTCG CCTAAAACTG CTTGTACCAA TTGCTATTGT AAAAAGTGTT 2220
GCTTTCATTG CCAAGTTTGT TTCATAACAA AAGCCTTAGG CATCTCCTAT GGCAGGAAGA 2280
AGCGGAGACA GCGACGAAGA GCTCATCAGA ACAGTCAGAC TCATCAAGCT TCTCTATCAA 2340
AGCAGTAAGT AGTACATGTA ACGCAACCTA TACCAATAGT AGCAATAGTA GCATTAGTAG 2400
TAGCAATAAT AATAGCAATA GTTGTGTGGT CCATAGTAAT CATAGAATAT AGGAAAATAT 2460
TAAGACAAAG AAAAATAGAC AGGTTAATTG ATAGACTAAT AGAAAGAGCA GAAGACAGTG 2520
GCAATGAGAG TGAAGGAGAA ATATCAGCAC TTGTGGAGAT GGGGGTGGAG ATGGGGCACC 2580
ATGCTCCTTG GGATGTTGAT GATCTGTAGT GCTACAGAAA AATTGTGGGT CACAGTCTAT 2640
TATGGGGTAC CTGTGTGGAA GGAAGCAACC ACCACTCTAT TTTGTGCATC AGATGCTAAA 2700
GCATAGATCT TCAGACTTGG AGGAGGAGAT ATGAGGGACA ATTGGAGAAG TGAATTATAT 2760
AAATATAAAG TAGTAAAAAT TGAACCATTA GGAGTAGCAC CCACCAAGGC AAAGAGAAGA 2820
GTGGTGCAGA GAGAAAAAAG AGCAGTGGGA ATAGGAGCTT TGTTCCTTGG GTTCTTGGGA 2880
GCAGCAGGAA GCACTATGGG CGCAGCGTCA ATGACGCTGA CGGTACAGGC CAGACAATTA 2940
TTGTCTGGTA TAGTGCAGCA GCAGAACAAT TTGCTGAGGG CTATTGAGGC GCAACAGCAT 3000
CTGTTGCAAC TCACAGTCTG GGGCATCAAG CAGCTCCAGG CAAGAATCCT GGCTGTGGAA 3060
AGATACCTAA AGGATCAACA GCTCCTGGGG ATTTGGGGTT GCTCTGGAAA ACTCATTTGC 3120
ACCACTGCTG TGCCTTGGAA TGCTAGTTGG AGTAATAAAT CTCTGGAACA GATCTGGAAT 3180
CACACGACCT GGATGGAGTG GGACAGAGAA ATTAACAATT ACACAAGCTT AATACACTCC 3240
TTAATTGAAG AATCGCAAAA CCAGCAAGAA AAGAATGAAC AAGAATTATT GGAATTAGAT 3300
AAATGGGCAA GTTTGTGGAA TTGGTTTAAC ATAACAAATT GGCTGTGGTA TATAAAATTA 3360
TTCATAATGA TAGTAGGAGG CTTGGTAGGT TTAAGAATAG TTTTTGCTGT ACTTTCTATA 3420
GTGAATAGAG TTAGGCAGGG ATATTCACCA TTATCGTTTC AGACCCACCT CCCAACCCCG 3480
AGGGGACCCG ACAGGCCCGA AGGAATAGAA GAAGAAGGTG GAGAGAGAGA CAGAGACAGA 3540
TCCATTCGAT TAGTGAACGG ATCCTTGGCA CTTATCTGGG ACGATCTGCG GAGCCTGTGC 3600
CTCTTCAGCT ACCACCGCTT GAGAGACTTA CTCTTGATTG TAACGAGGAT TGTGGAACTT 3660
CTGGGACGCA GGGGGTGGGA AGCCCTCAAA TATTGGTGGA ATCTCCTACA GTATTGGAGT 3720
CAGGAACTAA AGAATAGTGC TGTTAGCTTG CTCAATGCCA CAGCCATAGC AGTAGCTGAG 3780
GGGACAGATA GGGTTATAGA AGTAGTACAA GGAGCTTGTA GAGCTATTCG CCACATACCT 3840
AGAAGAATAA GACAGGGCTT GGAAAGGATT TTGCTATAAG ATGGGTGGCA AGTGGTCAAA 3900
AAGTAGTGTG ATTGGATGGC CTACTGTAAG GGAAAGAATG AGACGAGCTG AGCCAGCAGC 3960
AGATAGGGTG GGAGCAGCAT CTCGACGCTG CAGGAGTGGG GAGGCACGAT GGCCGCTTTG 4020
GTCGAGGCGG ATCCGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT AAATCAATAT 4080
TGGCTATTGG CCATTGCATA CGTTGTATCC ATATCATAAT ATGTACATTT ATATTGGCTC 4140
ATGTCCAACA TTACCGCCAT GTTGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT 4200
TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC TTACGGTAAA 4260
TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA TGACGTATGT 4320
TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT ATTTACGGTA 4380

AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT 4440
CAATGACGGT AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAT GGGACTTTCC 4500
TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC GGTTTTGGCA 4560
GTACATCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT 4620
TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA AATGTCGTAA 4680
CAACTCCGCC CCATTGACGC AAATGGGCGG TAGGCATGTA CGGTGGGAGG TCTATATAAG 4740
CAGAGCTCGT TTAGTGAACC GTCAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT 4800
CCATAGAAGA CACCGGGACC GATCCAGCCT CCGCGGCCCC AAGCTTCAGC TGCTCGAGCC 4860
CGGGGATGAC GTCATCGACT TCGAAGGTTC GAATCCTTCT ACTGCCACCA TTTTTTCTCT 4920
ACGTCATCGA CTTCGAAGGT TCGAATCCTT CCCTGTCCAC CAGTCGAGTA TTACGTCATC 4980
GACTTCGAAG GTTCGAATCC TTCTAGATTC ACCATTTTTT AGGAACGTCA TCGACTTCGA 5040
AGGTTCGAAT CCTTCCAGTT CCACCAGTCG AGGCCAACGT CATCGACTTC GAAGGTTCGA 5100
ATCCTTCTCT TCCCACCATT TTTTTTCCAC GTCATCGACT TCGAAGGTTC GAATCCTTCG 5160
GGGCCCACCA GTCGAGGGCT ACGTCATCGA CTTCGAAGGT TCGAATCCTT CTTGCTTCAC 5220
CATTTTTTCT GAACGTCATC GACTTCGAAG GTTCGAATCC TTCTGCTGTC ACCAGTCGAG 5280
TATAACGTCA TCGACTTCGA AGGTTCGAAT CCTTCACCGG TCACCATTTT TTTATAACGT 5340
CATCGACTTC GAAGGTTCGA ATCCTTCTTC TTACACCAGT CGAGGTACAC GTCATCGACT 5400
TCGAAGGTTC GAATCCTTCG TAGTTCACCA TTTTTTGTGC ACGTCATCGA CTTCGAAGGT 5460
TCGAATCCTT CTAGGCCCAC CAGTCGACGC ATGCCTGCAG GTCGAGGTCG ATACCGTCGA 5520
GACCTAGAAA AACATGGAGC AATCACAAGT AGCAATACAG CAGCTACCAA TGCTGATTGT 5580
GCCTGGCTAG AAGCACAAGA GGAGGAGGAG GTGGGTTTTC CAGTCACACC TCAGGTACCT 5640
TTAAGACCAA TGACTTACAA GGCAGCTGTA GATCTTAGCC ACTTTTTAAA AGAAAAGGGG 5700
GGACTGGAAG GGCTAATTCA CTCCCAACGA AGACAAGATA TCCTTGATCT GTGGATCTAC 5760
CACACACAAG GCTACTTCCC TGATTGGCAG AACTACACAC CAGGGCCAGG GATCAGATAT 5820
CCACTGACCT TTGGATGGTG CTACAAGCTA GTACCAGTTG AGCAAGAGAA GGTAGAAGAA 5880
GCCAATGAAG GAGAGAACAC CCGCTTGTTA CACCCTGTGA GCCTGCATGG GATGGATGAC 5940
CCGGAGAGAG AAGTATTAGA GTGGAGGTTT GACAGCCGCC TAGCATTTCA TCACATGGCC 6000
CGAGAGCTGC ATCCGGAGTA CTTCAAGAAC TGCTGACATC GAGCTTGCTA CAAGGGACTT 6060
TCCGCTGGGG ACTTTCCAGG GAGGCGTGGC CTGGGCGGGA CTGGGGAGTG GCGAGCCCTC 6120
AGATGCTGCA TATAAGCAGC TGCTTTTTGC CTGTACTGGG TCTCTCTGGT TAGACCAGAT 6180
CTGAGCCTGG GAGCTCTCTG GCTAACTAGG GAACCCACTG CTTAAGCCTC AATAAAGCTT 6240
GCCTTGAGTG CTTCAAGTAG TGTGTGCCCG TCTGTTGTGT GACTCTGGTA ACTAGAGATC 6300
CCTCAGACCC TTTTAGTCAG TGTGGAAAAT CTCTAGCAGT CGAGGGGGGG CCCGGTACCC 6360
AGCTTTTGTT CCCTTTAGTG AGGGTTAATT GCGCGCTTGG CGTAATCATG GTCATAGCTG 6420
TTTCCTGTGT GAAATTGTTA TCCGCTCACA ATTCCACACA ACATACGAGC CGGAAGCATA 6480
AAGTGTAAAG CCTGGGGTGC CTAATGAGTG AGCTAACTCA CATTAATTGC GTTGCGCTCA 6540
CTGCCCGCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT CGGCCAACGC 6600
GCGGGGAGAG GCGGTTTGCG TATTGGGCGC TCTTCCGCTT CCTCGCTCAC TGACTCGCTG 6660
CGCTCGGTCG TTCGGCTGCG GCGAGCGGTA TCAGCTCACT CAAAGGCGGT AATACGGTTA 6720
TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA GCAAAAGGCC 6780
AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTTCCATA GGCTCCGCCC CCCTGACGAG 6840
CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC 6900
CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCT GCCGCTTACC 6960
GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC TTTCTCATAG CTCACGCTGT 7020
AGGTATCTCA GTTCGGTGTA GGTCGTTCGC TCCAAGCTGG GCTGTGTGCA CGAACCCCCC 7080
GTTCAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGTAAGA 7140
CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA 7200
GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA 7260
TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA 7320
TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG TTTGCAAGCA GCAGATTACG 7380
CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTTT CTACGGGGTC TGACGCTCAG 7440
TGGAACGAAA ACTCACGTTA AGGGATTTTG GTCATGAGAT TATCAAAAAG GATCTTCACC 7500
TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT 7560
TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT 7620
CGTTCATCCA TAGTTGCCTG ACTCCCCGTC GTGTAGATAA CTACGATACG GGAGGGCTTA 7680
CCATCTGGCC CCAGTGCTGC AATGATACCG CGAGACCCAC GCTCACCGGC TCCAGATTTA 7740
TCAGCAATAA ACCAGCCAGC CGGAAGGGCC GAGCGCAGAA GTGGTCCTGC AACTTTATCC 7800
GCCTCCATCC AGTCTATTAA TTGTTGCCGG GAAGCTAGAG TAAGTAGTTC GCCAGTTAAT 7860
AGTTTGCGCA ACGTTGTTGC CATTGCTACA GGCATCGTGG TGTCACGCTC GTCGTTTGGT 7920
ATGGCTTCAT TCAGCTCCGG TTCCCAACGA TCAAGGCGAG TTACATGATC CCCCATGTTG 7980
TGCAAAAAAG CGGTTAGCTC CTTCGGTCCT CCGATCGTTG TCAGAAGTAA GTTGGCCGCA 8040
GTGTTATCAC TCATGGTTAT GGCAGCACTG CATAATTCTC TTACTGTCAT GCCATCCGTA 8100
AGATGCTTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT TCTGAGAATA GTGTATGCGG 8160
CGACCGAGTT GCTCTTGCCC GGCGTCAATA CGGGATAATA CCGCGCCACA TAGCAGAACT 8220
TTAAAAGTGC TCATCATTGG AAAACGTTCT TCGGGGCGAA AACTCTCAAG GATCTTACCG 8280

```
CTGTTGAGAT CCAGTTCGAT GTAACCCACT CGTGCACCCA ACTGATCTTC AGCATCTTTT 8340
ACTTTCACCA GCGTTTCTGG GTGAGCAAAA ACAGGAAGGC AAAATGCCGC AAAAAAGGGA 8400
ATAAGGGCGA CACGGAAATG TTGAATACTC ATACTCTTCC TTTTTCAATA TTATTGAAGC 8460
ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG AATGTATTTA GAAAAATAAA 8520
CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC                       8560
```

SEQ. I.D. NO. 12 – pSYNGP2 – codon optimised HIV-1 gagpol with leader sequence

```
   1 GGGTCTCTCT GGTTAGACCA GATCTGAGCC TGGGAGCTCT CTGGCTAACT AGGGAACCCA
  61 CTGCTTAAGC CTCAATAAAG CTTGCCTTGA GTGCTTCAAG TAGTGTGTGC CCGTCTGTTG
 121 TGTGACTCTG GTAACTAGAG ATCCCTCAGA CCCTTTTAGT CAGTGTGGAA AATCTCTAGC
 181 AGTGGCGCCC GAACAGGGAC CTGAAAGCGA AAGGGAAACC AGAGCTCTCT CGACGCAGGA
 241 CTCGGCTTGC TGAAGCGCCC GCACGGCAAG AGGCGAGGGG CGGCGACTGG TGAGTACGCC
 301 AAAAATTTTG ACTAGCGGAG GCTAGAAGGA GAGAGATGGG CGCCCGCGCC AGCGTGCTGT
 361 CGGGCGGCGA GCTGGACCGC TGGGAGAAGA TCCGCCTGCG CCCCGGCGGC AAAAAGAAGT
 421 ACAAGCTGAA GCACATCGTG TGGGCCAGCC GCGAACTGGA GCGCTTCGCC GTGAACCCCG
 481 GGCTCCTGGA GACCAGCGAG GGGTGCCGCC AGATCCTCGG CCAACTGCAG CCCAGCCTGC
 541 AAACCGGCAG CGAGGAGCTG CGCAGCCTGT ACAACACCGT GGCCACGCTG TACTGCGTCC
 601 ACCAGCGCAT CGAAATCAAG GATACGAAAG AGGCCCTGGA TAAAATCGAA GAGGAACAGA
 661 ATAAGAGCAA AAAGAAGGCC CAACAGGCCG CCGCGGACAC CGGACACAGC AACCAGGTCA
 721 GCCAGAACTA CCCCATCGTG CAGAACATCC AGGGGCAGAT GGTGCACCAG GCCATCTCCC
 781 CCCGCACGCT GAACGCCTGG GTGAAGGTGG TGGAAGAGAA GGCTTTTAGC CCGGAGGTGA
 841 TACCCATGTT CTCAGCCCTG TCAGAGGGAG CCACCCCCCA AGATCTGAAC ACCATGCTCA
 901 ACACAGTGGG GGGACACCAG GCCGCCATGC AGATGCTGAA GGAGACCATC AATGAGGAGG
 961 CTGCCGAATG GGATCGTGTG CATCCGGTGC ACGCAGGGCC CATCGCACCG GGCCAGATGC
1021 GTGAGCCACG GGGCTCAGAC ATCGCCGGAA CGACTAGTAC CCTTCAGGAA CAGATCGGCT
1081 GGATGACCAA CAACCCACCC ATCCCGGTGG GAGAAATCTA CAAACGCTGG ATCATCCTGG
1141 GCCTGAACAA GATCGTGCGC ATGTATAGCC CTACCAGCAT CCTGGACATC CGCCAAGGCC
1201 CGAAGGAACC CTTTCGCGAC TACGTGGACC GGTTCTACAA AACGCTCCGC GCCGAGCAGG
1261 CTAGCCAGGA GGTGAAGAAC TGGATGACCG AAACCCTGCT GGTCCAGAAC GCGAACCCGG
1321 ACTGCAAGAC GATCCTGAAG GCCCTGGGCC CAGCGGCTAC CCTAGAGGAA ATGATGACCG
1381 CCTGTCAGGG AGTGGGCGGA CCCGGCCACA AGGCACGCGT CCTGGCTGAG GCCATGAGCC
1441 AGGTGACCAA CTCCGCTACC ATCATGATGC AGCGCGGCAA CTTTCGGAAC CAACGCAAGA
1501 TCGTCAAGTG CTTCAACTGT GGCAAAGAAG GGCACACAGC CCGCAACTGC AGGGCCCCTA
1561 GGAAAAAGGG CTGTTGGAAA TGTGGAAAGG AAGGACACCA AATGAAAGAT TGTACTGAGA
1621 GACAGGCTAA TTTTTTAGGG AAGATCTGGC CTTCCCACAA GGGAAGGCCA GGGAATTTTC
1681 TTCAGAGCAG ACCAGAGCCA ACAGCCCCAC CAGAAGAGAG CTTCAGGTTT GGGGAAGAGA
1741 CAACAACTCC CTCTCAGAAG CAGGAGCCGA TAGACAAGGA ACTGTATCCT TTAGCTTCCC
1801 TCAGATCACT CTTTGGCAGC GACCCCTCGT CACAATAAAG ATAGGGGGGC AGCTCAAGGA
1861 GGCTCTCCTG GACACCGGAG CAGACGACAC CGTGCTGGAG GAGATGTCGT TGCCAGGCCG
1921 CTGGAAGCCG AAGATGATCG GGGGAATCGG CGGTTTCATC AAGGTGCGCC AGTATGACCA
1981 GATCCTCATC GAAATCTGCG GCCACAAGGC TATCGGTACC GTGCTGGTGG GCCCCACACC
2041 CGTCAACATC ATCGGACGCA ACCTGTTGAC GCAGATCGGT TGCACGCTGA ACTTCCCCAT
2101 TAGCCCTATC GAGACGGTAC CGGTGAAGCT GAAGCCCGGG ATGGACGGCC CGAAGGTCAA
2161 GCAATGGCCA TTGACAGAGG AGAAGATCAA GGCACTGGTG GAGATTTGCA CAGAGATGGA
2221 AAAGGAAGGG AAAATCTCCA AGATTGGGCC TGAGAACCCG TACAACACGC CGGTGTTCGC
2281 AATCAAGAAG AAGGACTCGA CGAAATGGCG CAAGCTGGTG GACTTCCGCG AGCTGAACAA
2341 GCGCACGCAA GACTTCTGGG AGGTTCAGCT GGGCATCCCG CACCCCGCAG GGCTGAAGAA
2401 GAAGAAATCC GTGACCGTAC TGGATGTGGG TGATGCCTAC TTCTCCGTTC CCCTGGACGA
2461 AGACTTCAGG AAGTACACTG CCTTCACAAT CCCTTCGATC AACAACGAGA CACCGGGGAT
2521 TCGATATCAG TACAACGTGC TGCCCCAGGG CTGGAAAGGC TCTCCCGCAA TCTTCCAGAG
2581 TAGCATGACC AAAATCCTGG AGCCTTTCCG CAAACAGAAC CCCGACATCG TCATCTATCA
2641 GTACATGGAT GACTTGTACG TGGGCTCTGA TCTAGAGATA GGGCAGCACC GCACCAAGAT
2701 CGAGGAGCTG CGCCAGCACC TGTTGAGGTG GGGACTGACC ACACCCGACA AGAAGCACCA
2761 GAAGGAGCCT CCCTTCCTCT GGATGGGTTA CGAGCTGCAC CCTGACAAAT GGACCGTGCA
2821 GCCTATCGTG CTGCCAGAGA AAGACAGCTG GACTGTCAAC GACATACAGA AGCTGGTGGG
2881 GAAGTTGAAC TGGGCCAGTC AGATTTACCC AGGGATTAAG GTGAGGCAGC TGTGCAAACT
2941 CCTCCGCGGA ACCAAGGCAC TCACAGAGGT GATCCCCCTA ACCGAGGAGG CCGAGCTCGA
3001 ACTGGCAGAA AACCGAGAGA TCCTAAAGGA GCCCGTGCAC GGCGTGTACT ATGACCCCTC
3061 CAAGGACCTG ATCGCCGAGA TCCAGAAGCA GGGGCAAGGC CAGTGGACCT ATCAGATTTA
3121 CCAGGAGCCC TTCAAGAACC TGAAGACCGG CAAGTACGCC CGGATGAGGG GTGCCCACAC
3181 TAACGACGTC AAGCAGCTGA CCGAGGCCGT GCAGAAGATC ACCACCGAAA GCATCGTGAT
3241 CTGGGGAAAG ACTCCTAAGT TCAAGCTGCC CATCCAGAAG GAAACCTGGG AAACCTGGTG
3301 GACAGAGTAT TGGCAGGCCA CCTGGATTCC TGAGTGGGAG TTCGTCAACA CCCCTCCCCT
3361 GGTGAAGCTG TGGTACCAGC TGGAGAAGGA GCCCATAGTG GGCGCCGAAA CCTTCTACGT
```

```
3421 GGATGGGGCC GCTAACAGGG AGACTAAGCT GGGCAAAGCC GGATACGTCA CTAACCGGGG
3481 CAGACAGAAG GTTGTCACCC TCACTGACAC CACCAACCAG AAGACTGAGC TGCAGGCCAT
3541 TTACCTCGCT TTGCAGGACT CGGGCCTGGA GGTGAACATC GTGACAGACT CTCAGTATGC
3601 CCTGGGCATC ATTCAAGCCC AGCCAGACCA GAGTGAGTCC GAGCTGGTCA ATCAGATCAT
3661 CGAGCAGCTG ATCAAGAAGG AAAAGGTCTA TCTGGCCTGG GTACCCGCCC ACAAAGGCAT
3721 TGGCGGCAAT GAGCAGGTCG ACAAGCTGGT CTCGGCTGGC ATCAGGAAGG TGCTATTCCT
3781 GGATGGCATC GACAAGGCCC AGGACGAGCA CGAGAAATAC CACAGCAACT GGCGGGCCAT
3841 GGCTAGCGAC TTCAACCTGC CCCCTGTGGT GGCCAAAGAG ATCGTGGCCA GCTGTGACAA
3901 GTGTCAGCTC AAGGGCGAAG CCATGCATGG CCAGGTGGAC TGTAGCCCCG GCATCTGGCA
3961 ACTCGATTGC ACCCATCTGG AGGGCAAGGT TATCCTGGTA GCCGTCCATG TGGCCAGTGG
4021 CTACATCGAG GCCGAGGTCA TTCCCGCCGA AACAGGGCAG GAGACAGCCT ACTTCCTCCT
4081 GAAGCTGGCA GGCCGGTGGC CAGTGAAGAC CATCCATACT GACAATGGCA GCAATTTCAC
4141 CAGTGCTACG GTTAAGGCCG CCTGCTGGTG GGCGGGAATC AAGCAGGAGT TCGGGATCCC
4201 CTACAATCCC CAGAGTCAGG GCGTCGTCGA GTCTATGAAT AAGGAGTTAA AGAAGATTAT
4261 CGGCCAGGTC AGAGATCAGG CTGAGCATCT CAAGACCGCG GTCCAAATGG CGGTATTCAT
4321 CCACAATTTC AAGCGGAAGG GGGGGATTGG GGGGTACAGT GCGGGGGAGC GGATCGTGGA
4381 CATCATCGCG ACCGACATCC AGACTAAGGA GCTGCAAAAG CAGATTACCA AGATTCAGAA
4441 TTTCCGGGTC TACTACAGGG ACAGCAGAAA TCCCCTCTGG AAAGGCCCAG CGAAGCTCCT
4501 CTGGAAGGGT GAGGGGGCAG TAGTGATCCA GGATAATAGC GACATCAAGG TGGTGCCCAG
4561 AAGAAAGGCG AAGATCATTA GGGATTATGG CAAACAGATG GCGGGTGATG ATTGCGTGGC
4621 GAGCAGACAG GATGAGGATT AG
```

SEQ. I.D. NO. 13 – pSYNGP3 – codon optimised HIV-1 gagpol with leader sequence from the major splice donor

```
   1 GTGAGTACGC CAAAAATTTT GACTAGCGGA GGCTAGAAGG AGAGAGATGG GCGCCCGCGC
  61 CAGCGTGCTG TCGGGCGGCG AGCTGGACCG CTGGGAGAAG ATCCGCCTGC GCCCCGGCGG
 121 CAAAAAGAAG TACAAGCTGA AGCACATCGT GTGGGCCAGC CGCGAACTGG AGCGCTTCGC
 181 CGTGAACCCC GGGCTCCTGG AGACCAGCGA GGGGTGCCGC CAGATCCTCG GCCAACTGCA
 241 GCCCAGCCTG CAAACCGGCA GCGAGGAGCT GCGCAGCCTG TACAACACCG TGGCCACGCT
 301 GTACTGCGTC CACCAGCGCA TCGAAATCAA GGATACGAAA GAGGCCCTGG ATAAAATCGA
 361 AGAGGAACAG AATAAGAGCA AAAAGAAGGC CCAACAGGCC GCCGCGGACA CCGGACACAG
 421 CAACCAGGTC AGCCAGAACT ACCCCATCGT GCAGAACATC CAGGGGCAGA TGGTGCACCA
 481 GGCCATCTCC CCCCGCACGC TGAACGCCTG GGTGAAGGTG GTGGAAGAGA AGGCTTTTAG
 541 CCCGGAGGTG ATACCCATGT TCTCAGCCCT GTCAGAGGGA GCCACCCCCC AAGATCTGAA
 601 CACCATGCTC AACACAGTGG GGGGACACCA GGCCGCCATG CAGATGCTGA AGGAGACCAT
 661 CAATGAGGAG GCTGCCGAAT GGGATCGTGT GCATCCGGTG CACGCAGGGC CCATCGCACC
 721 GGGCCAGATG CGTGAGCCAC GGGGCTCAGA CATCGCCGGA ACGACTAGTA CCCTTCAGGA
 781 ACAGATCGGC TGGATGACCA ACAACCCACC CATCCCGGTG GGAGAAATCT ACAAACGCTG
 841 GATCATCCTG GGCCTGAACA AGATCGTGCG CATGTATAGC CCTACCAGCA TCCTGGACAT
 901 CCGCCAAGGC CCGAAGGAAC CCTTTCGCGA CTACGTGGAC CGGTTCTACA AAACGCTCCG
 961 CGCCGAGCAG GCTAGCCAGG AGGTGAAGAA CTGGATGACC GAAACCCTGC TGGTCCAGAA
1021 CGCGAACCCG GACTGCAAGA CGATCCTGAA GGCCCTGGGC CCAGCGGCTA CCCTAGAGGA
1081 AATGATGACC GCCTGTCAGG GAGTGGGCGG ACCCGGCCAC AAGGCACGCG TCCTGGCTGA
1141 GGCCATGAGC CAGGTGACCA ACTCCGCTAC CATCATGATG CAGCGCGGCA ACTTTCGGAA
1201 CCAACGCAAG ATCGTCAAGT GCTTCAACTG TGGCAAAGAA GGGCACACAG CCCGCAACTG
1261 CAGGGCCCCT AGGAAAAAGG GCTGTTGGAA ATGTGGAAAG GAAGGACACC AAATGAAAGA
1321 TTGTACTGAG AGACAGGCTA ATTTTTTAGG GAAGATCTGG CCTTCCCACA AGGGAAGGCC
1381 AGGGAATTTT CTTCAGAGCA GACCAGAGCC AACAGCCCCA CCAGAAGAGA GCTTCAGGTT
1441 TGGGGAAGAG ACAACAACTC CCTCTCAGAA GCAGGAGCCG ATAGACAAGG AACTGTATCC
1501 TTTAGCTTCC CTCAGATCAC TCTTTGGCAG CGACCCCTCG TCACAATAAA GATAGGGGGG
1561 CAGCTCAAGG AGGCTCTCCT GGACACCGGA GCAGACGACA CCGTGCTGGA GGAGATGTCG
1621 TTGCCAGGCC GCTGGAAGCC GAAGATGATC GGGGGAATCG GCGGTTTCAT CAAGGTGCGC
1681 CAGTATGACC AGATCCTCAT CGAAATCTGC GGCCACAAGG CTATCGGTAC CGTGCTGGTG
1741 GGCCCCACAC CCGTCAACAT CATCGGACGC AACCTGTTGA CGCAGATCGG TTGCACGCTG
1801 AACTTCCCCA TTAGCCCTAT CGAGACGGTA CCGGTGAAGC TGAAGCCCGG GATGGACGGC
1861 CCGAAGGTCA AGCAATGGCC ATTGACAGAG GAGAAGATCA AGGCACTGGT GGAGATTTGC
1921 ACAGAGATGG AAAAGGAAGG GAAAATCTCC AAGATTGGGC CTGAGAACCC GTACAACACG
1981 CCGGTGTTCG CAATCAAGAA GAAGGACTCG ACGAAATGGC GCAAGCTGGT GGACTTCCGC
2041 GAGCTGAACA AGCGCACGCA AGACTTCTGG GAGGTTCAGC TGGGCATCCC GCACCCCGCA
2101 GGGCTGAAGA AGAAGAAATC CGTGACCGTA CTGGATGTGG GTGATGCCTA CTTCTCCGTT
2161 CCCCTGGACG AAGACTTCAG GAAGTACACT GCCTTCACAA TCCCTTCGAT CAACAACGAG
2221 ACACCGGGGA TTCGATATCA GTACAACGTG CTGCCCCAGG GCTGGAAAGG CTCTCCCGCA
```

```
2281 ATCTTCCAGA GTAGCATGAC CAAAATCCTG GAGCCTTTCC GCAAACAGAA CCCCGACATC
2341 GTCATCTATC AGTACATGGA TGACTTGTAC GTGGGCTCTG ATCTAGAGAT AGGGCAGCAC
2401 CGCACCAAGA TCGAGGAGCT GCGCCAGCAC CTGTTGAGGT GGGGACTGAC CACACCCGAC
2461 AAGAAGCACC AGAAGGAGCC TCCCTTCCTC TGGATGGGTT ACGAGCTGCA CCCTGACAAA
2521 TGGACCGTGC AGCCTATCGT GCTGCCAGAG AAAGACAGCT GGACTGTCAA CGACATACAG
2581 AAGCTGGTGG GGAAGTTGAA CTGGGCCAGT CAGATTTACC CAGGGATTAA GGTGAGGCAG
2641 CTGTGCAAAC TCCTCCGCGG AACCAAGGCA CTCACAGAGG TGATCCCCCT AACCGAGGAG
2701 GCCGAGCTCG AACTGGCAGA AAACCGAGAG ATCCTAAAGG AGCCCGTGCA CGGCGTGTAC
2761 TATGACCCCT CCAAGGACCT GATCGCCGAG ATCCAGAAGC AGGGGCAAGG CCAGTGGACC
2821 TATCAGATTT ACCAGGAGCC CTTCAAGAAC CTGAAGACCG GCAAGTACGC CCGGATGAGG
2881 GGTGCCCACA CTAACGACGT CAAGCAGCTG ACCGAGGCCG TGCAGAAGAT CACCACCGAA
2941 AGCATCGTGA TCTGGGGAAA GACTCCTAAG TTCAAGCTGC CCATCCAGAA GGAAACCTGG
3001 GAAACCTGGT GGACAGAGTA TTGGCAGGCC ACCTGGATTC CTGAGTGGGA GTTCGTCAAC
3061 ACCCCTCCCC TGGTGAAGCT GTGGTACCAG CTGGAGAAGG AGCCCATAGT GGGCGCCGAA
3121 ACCTTCTACG TGGATGGGGC CGCTAACAGG GAGACTAAGC TGGGCAAAGC CGGATACGTC
3181 ACTAACCGGG GCAGACAGAA GGTTGTCACC CTCACTGACA CCACCAACCA GAAGACTGAG
3241 CTGCAGGCCA TTTACCTCGC TTTGCAGGAC TCGGGCCTGG AGGTGAACAT CGTGACAGAC
3301 TCTCAGTATG CCCTGGGCAT CATTCAAGCC CAGCCAGACC AGAGTGAGTC CGAGCTGGTC
3361 AATCAGATCA TCGAGCAGCT GATCAAGAAG GAAAAGGTCT ATCTGGCCTG GGTACCCGCC
3421 CACAAAGGCA TTGGCGGCAA TGAGCAGGTC GACAAGCTGG TCTCGGCTGG CATCAGGAAG
3481 GTGCTATTCC TGGATGGCAT CGACAAGGCC CAGGACGAGC ACGAGAAATA CCACAGCAAC
3541 TGGCGGGCCA TGGCTAGCGA CTTCAACCTG CCCCCTGTGG TGGCCAAAGA GATCGTGGCC
3601 AGCTGTGACA AGTGTCAGCT CAAGGGCGAA GCCATGCATG GCCAGGTGGA CTGTAGCCCC
3661 GGCATCTGGC AACTCGATTG CACCCATCTG GAGGGCAAGG TTATCCTGGT AGCCGTCCAT
3721 GTGGCCAGTG GCTACATCGA GGCCGAGGTC ATTCCCGCCG AAACAGGGCA GGAGACAGCC
3781 TACTTCCTCC TGAAGCTGGC AGGCCGGTGG CCAGTGAAGA CCATCCATAC TGACAATGGC
3841 AGCAATTTCA CCAGTGCTAC GGTTAAGGCC GCCTGCTGGT GGGCGGGAAT CAAGCAGGAG
3901 TTCGGGATCC CCTACAATCC CCAGAGTCAG GGCGTCGTCG AGTCTATGAA TAAGGAGTTA
3961 AAGAAGATTA TCGGCCAGGT CAGAGATCAG GCTGAGCATC TCAAGACCGC GGTCCAAATG
4021 GCGGTATTCA TCCACAATTT CAAGCGGAAG GGGGGGATTG GGGGGTACAG TGCGGGGGAG
4081 CGGATCGTGG ACATCATCGC GACCGACATC CAGACTAAGG AGCTGCAAAA GCAGATTACC
4141 AAGATTCAGA ATTTCCGGGT CTACTACAGG GACAGCAGAA ATCCCCTCTG GAAAGGCCCA
4201 GCGAAGCTCC TCTGGAAGGG TGAGGGGGCA GTAGTGATCC AGGATAATAG CGACATCAAG
4261 GTGGTGCCCA GAAGAAAGGC GAAGATCATT AGGGATTATG GCAAACAGAT GGCGGGTGAT
4321 GATTGCGTGG CGAGCAGACA GGATGAGGAT TAG
```

SEQ. I.D. NO. 14 – pSYNGP4 – codon optimised HIV-1 gagpol with 20 bp of the leader sequence of HIV-1, upstream of the start codon of ATG.

```
   1 CGGAGGCTAG AAGGAGAGAG ATGGGCGCCC GCGCCAGCGT GCTGTCGGGC GGCGAGCTGG
  61 ACCGCTGGGA GAAGATCCGC CTGCGCCCCG GCGGCAAAAA GAAGTACAAG CTGAAGCACA
 121 TCGTGTGGGC CAGCCGCGAA CTGGAGCGCT TCGCCGTGAA CCCCGGGCTC CTGGAGACCA
 181 GCGAGGGGTG CCGCCAGATC CTCGGCCAAC TGCAGCCCAG CCTGCAAACC GGCAGCGAGG
 241 AGCTGCGCAG CCTGTACAAC ACCGTGGCCA CGCTGTACTG CGTCCACCAG CGCATCGAAA
 301 TCAAGGATAC GAAAGAGGCC CTGGATAAAA TCGAAGAGGA ACAGAATAAG AGCAAAAAGA
 361 AGGCCCAACA GGCCGCCGCG GACACCGGAC ACAGCAACCA GGTCAGCCAG AACTACCCCA
 421 TCGTGCAGAA CATCCAGGGG CAGATGGTGC ACCAGGCCAT CTCCCCCCGC ACGCTGAACG
 481 CCTGGGTGAA GGTGGTGGAA GAGAAGGCTT TTAGCCCGGA GGTGATACCC ATGTTCTCAG
 541 CCCTGTCAGA GGGAGCCACC CCCCAAGATC TGAACACCAT GCTCAACACA GTGGGGGGAC
 601 ACCAGGCCGC CATGCAGATG CTGAAGGAGA CCATCAATGA GGAGGCTGCC GAATGGGATC
 661 GTGTGCATCC GGTGCACGCA GGGCCCATCG CACCGGGCCA GATGCGTGAG CCACGGGGCT
 721 CAGACATCGC CGGAACGACT AGTACCCTTC AGGAACAGAT CGGCTGGATG ACCAACAACC
 781 CACCCATCCC GGTGGGAGAA ATCTACAAAC GCTGGATCAT CCTGGGCCTG AACAAGATCG
 841 TGCGCATGTA TAGCCCTACC AGCATCCTGG ACATCCGCCA AGGCCCGAAG GAACCCTTTC
 901 GCGACTACGT GGACCGGTTC TACAAAACGC TCCGCGCCGA GCAGGCTAGC CAGGAGGTGA
 961 AGAACTGGAT GACCGAAACC CTGCTGGTCC AGAACGCGAA CCCGGACTGC AAGACGATCC
1021 TGAAGGCCCT GGGCCCAGCG GCTACCCTAG AGGAAATGAT GACCGCCTGT CAGGGAGTGG
1081 GCGGACCCGG CCACAAGGCA CGCGTCCTGG CTGAGGCCAT GAGCCAGGTG ACCAACTCCG
1141 CTACCATCAT GATGCAGCGC GGCAACTTTC GGAACCAACG CAAGATCGTC AAGTGCTTCA
1201 ACTGTGGCAA AGAAGGGCAC ACAGCCCGCA ACTGCAGGGC CCCTAGGAAA AAGGGCTGTT
1261 GGAAATGTGG AAAGGAAGGA CACCAAATGA AAGATTGTAC TGAGAGACAG GCTAATTTTT
1321 TAGGGAAGAT CTGGCCTTCC CACAAGGGAA GGCCAGGGAA TTTTCTTCAG AGCAGACCAG
1381 AGCCAACAGC CCCACCAGAA GAGAGCTTCA GGTTTGGGGA AGAGACAACA ACTCCCTCTC
1441 AGAAGCAGGA GCCGATAGAC AAGGAACTGT ATCCTTTAGC TTCCCTCAGA TCACTCTTTG
```

```
1501 GCAGCGACCC CTCGTCACAA TAAAGATAGG GGGGCAGCTC AAGGAGGCTC TCCTGGACAC
1561 CGGAGCAGAC GACACCGTGC TGGAGGAGAT GTCGTTGCCA GGCCGCTGGA AGCCGAAGAT
1621 GATCGGGGGA ATCGGCGGTT TCATCAAGGT GCGCCAGTAT GACCAGATCC TCATCGAAAT
1681 CTGCGGCCAC AAGGCTATCG GTACCGTGCT GGTGGGCCCC ACACCCGTCA ACATCATCGG
1741 ACGCAACCTG TTGACGCAGA TCGGTTGCAC GCTGAACTTC CCCATTAGCC CTATCGAGAC
1801 GGTACCGGTG AAGCTGAAGC CCGGGATGGA CGGCCCGAAG GTCAAGCAAT GGCCATTGAC
1861 AGAGGAGAAG ATCAAGGCAC TGGTGGAGAT TTGCACAGAG ATGGAAAAGG AAGGGAAAAT
1921 CTCCAAGATT GGGCCTGAGA ACCCGTACAA CACGCCGGTG TTCGCAATCA AGAAGAAGGA
1981 CTCGACGAAA TGGCGCAAGC TGGTGGACTT CCGCGAGCTG AACAAGCGCA CGCAAGACTT
2041 CTGGGAGGTT CAGCTGGGCA TCCCGCACCC CGCAGGGCTG AAGAAGAAGA AATCCGTGAC
2101 CGTACTGGAT GTGGGTGATG CCTACTTCTC CGTTCCCCTG GACGAAGACT TCAGGAAGTA
2161 CACTGCCTTC ACAATCCCTT CGATCAACAA CGAGACACCG GGGATTCGAT ATCAGTACAA
2221 CGTGCTGCCC CAGGGCTGGA AAGGCTCTCC CGCAATCTTC CAGAGTAGCA TGACCAAAAT
2281 CCTGGAGCCT TTCCGCAAAC AGAACCCCGA CATCGTCATC TATCAGTACA TGGATGACTT
2341 GTACGTGGGC TCTGATCTAG AGATAGGGCA GCACCGCACC AAGATCGAGG AGCTGCGCCA
2401 GCACCTGTTG AGGTGGGGAC TGACCACACC CGACAAGAAG CACCAGAAGG AGCCTCCCTT
2461 CCTCTGGATG GGTTACGAGC TGCACCCTGA CAAATGGACC GTGCAGCCTA TCGTGCTGCC
2521 AGAGAAAGAC AGCTGGACTG TCAACGACAT ACAGAAGCTG GTGGGGAAGT TGAACTGGGC
2581 CAGTCAGATT TACCCAGGGA TTAAGGTGAG GCAGCTGTGC AAACTCCTCC GCGGAACCAA
2641 GGCACTCACA GAGGTGATCC CCCTAACCGA GGAGGCCGAG CTCGAACTGG CAGAAAACCG
2701 AGAGATCCTA AAGGAGCCCG TGCACGGCGT GTACTATGAC CCCTCCAAGG ACCTGATCGC
2761 CGAGATCCAG AAGCAGGGGC AAGGCCAGTG GACCTATCAG ATTTACCAGG AGCCCTTCAA
2821 GAACCTGAAG ACCGGCAAGT ACGCCCGGAT GAGGGGTGCC CACACTAACG ACGTCAAGCA
2881 GCTGACCGAG GCCGTGCAGA AGATCACCAC CGAAAGCATC GTGATCTGGG GAAAGACTCC
2941 TAAGTTCAAG CTGCCCATCC AGAAGGAAAC CTGGGAAACC TGGTGGACAG AGTATTGGCA
3001 GGCCACCTGG ATTCCTGAGT GGGAGTTCGT CAACACCCCT CCCCTGGTGA AGCTGTGGTA
3061 CCAGCTGGAG AAGGAGCCCA TAGTGGGCGC CGAAACCTTC TACGTGGATG GGGCCGCTAA
3121 CAGGGAGACT AAGCTGGGCA AAGCCGGATA CGTCACTAAC CGGGGCAGAC AGAAGGTTGT
3181 CACCCTCACT GACACCACCA ACCAGAAGAC TGAGCTGCAG GCCATTTACC TCGCTTTGCA
3241 GGACTCGGGC CTGGAGGTGA ACATCGTGAC AGACTCTCAG TATGCCCTGG GCATCATTCA
3301 AGCCCAGCCA GACCAGAGTG AGTCCGAGCT GGTCAATCAG ATCATCGAGC AGCTGATCAA
3361 GAAGGAAAAG GTCTATCTGG CCTGGGTACC CGCCCACAAA GGCATTGGCG GCAATGAGCA
3421 GGTCGACAAG CTGGTCTCGG CTGGCATCAG GAAGGTGCTA TTCCTGGATG GCATCGACAA
3481 GGCCCAGGAC GAGCACGAGA AATACCACAG CAACTGGCGG GCCATGGCTA GCGACTTCAA
3541 CCTGCCCCCT GTGGTGGCCA AAGAGATCGT GGCCAGCTGT GACAAGTGTC AGCTCAAGGG
3601 CGAAGCCATG CATGGCCAGG TGGACTGTAG CCCCGGCATC TGGCAACTCG ATTGCACCCA
3661 TCTGGAGGGC AAGGTTATCC TGGTAGCCGT CCATGTGGCC AGTGGCTACA TCGAGGCCGA
3721 GGTCATTCCC GCCGAAACAG GGCAGGAGAC AGCCTACTTC CTCCTGAAGC TGGCAGGCCG
3781 GTGGCCAGTG AAGACCATCC ATACTGACAA TGGCAGCAAT TTCACCAGTG CTACGGTTAA
3841 GGCCGCCTGC TGGTGGGCGG GAATCAAGCA GGAGTTCGGG ATCCCCTACA ATCCCCAGAG
3901 TCAGGGCGTC GTCGAGTCTA TGAATAAGGA GTTAAAGAAG ATTATCGGCC AGGTCAGAGA
3961 TCAGGCTGAG CATCTCAAGA CCGCGGTCCA AATGGCGGTA TTCATCCACA ATTTCAAGCG
4021 GAAGGGGGGG ATTGGGGGGT ACAGTGCGGG GGAGCGGATC GTGGACATCA TCGCGACCGA
4081 CATCCAGACT AAGGAGCTGC AAAAGCAGAT TACCAAGATT CAGAATTTCC GGGTCTACTA
4141 CAGGGACAGC AGAAATCCCC TCTGGAAAGG CCCAGCGAAG CTCCTCTGGA AGGGTGAGGG
4201 GGCAGTAGTG ATCCAGGATA ATAGCGACAT CAAGGTGGTG CCCAGAAGAA AGGCGAAGAT
4261 CATTAGGGAT TATGGCAAAC AGATGGCGGG TGATGATTGC GTGGCGAGCA GACAGGATGA
4321 GGATTAG
```

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 00/01002

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/86 C12N9/00 C12N9/22 C12N7/04 C12N5/10
//A61P31/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
MEDLINE, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 20060 A (UNIV JOHNS HOPKINS MED) 5 June 1997 (1997-06-05) page 9 -page 10 page 14, line 6 -page 17 page 20, line 30 -page 21 page 27, line 35 -page 32, line 3 examples	1-7,12, 13,16-23
Y	claims	8-11,14, 15
Y	WO 98 17815 A (MITROPHANOUS KYRIACOS ;KIM NARRY (GB); KINGSMAN ALAN J (GB); KINGS) 30 April 1998 (1998-04-30) the whole document	8-11

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 June 2000	10/07/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Andres, S

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	HAAS J ET AL: "CODON USAGE LIMITATION IN THE EXPRESSION OF HIV-1 ENVELOPE GLYCOPROTEIN" CURRENT BIOLOGY, vol. 6, no. 3, 1 March 1996 (1996-03-01), pages 315-324, XP000619599 ISSN: 0960-9822 cited in the application the whole document	14,15
A	EP 0 711 829 A (VIAGENE INC) 15 May 1996 (1996-05-15) page 21, line 20 - line 27	1-23
A	YUAN Y ET AL: "Targeted cleavage of mRNA by human RNase P" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 89, no. 89, September 1992 (1992-09), pages 8006-8010, XP002104826 ISSN: 0027-8424 cited in the application	
T	WO 99 41397 A (MITROPHANOUS KYRIACOS ;KINGSMAN ALAN JOHN (GB); OXFORD BIOMEDICAL) 19 August 1999 (1999-08-19) the whole document	1-23
T	KOTSOPOULOU E. ET AL.: "A Rev-independent human immunodeficiency virus type 1 (HIV-1)-based vector that exploits a codon-optimized HIV-1 gag-pol gene." J VIROL 2000 MAY;74(10):4839-52, XP002140792 the whole document	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/EP 00/01002

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9720060 A	05-06-1997	AU 1124997 A	19-06-1997
		CN 1207775 A	10-02-1999
		CZ 9801624 A	16-12-1998
		EP 0871757 A	21-10-1998
		JP 2000503527 T	28-03-2000
		NO 982418 A	27-07-1998
WO 9817815 A	30-04-1998	AU 4712297 A	15-05-1998
		AU 4712397 A	15-05-1998
		CN 1234836 A	10-11-1999
		CN 1234075 A	03-11-1999
		CZ 9901372 A	14-07-1999
		CZ 9901373 A	14-07-1999
		EP 0904392 A	31-03-1999
		EP 0939827 A	08-09-1999
		WO 9817816 A	30-04-1998
		GB 2325003 A,B	11-11-1998
		GB 2331522 A	26-05-1999
		NO 991742 A	13-04-1999
		NO 991781 A	17-06-1999
		PL 332875 A	25-10-1999
		PL 332876 A	25-10-1999
EP 0711829 A	15-05-1996	EP 0716148 A	12-06-1996
		EP 0814154 A	29-12-1997
		EP 0982405 A	01-03-2000
		AU 3682699 A	07-10-1999
		AU 703653 B	01-04-1999
		AU 5645098 A	04-06-1998
		AU 690583 B	30-04-1998
		AU 7835894 A	03-04-1995
		CA 2158937 A	23-03-1995
		EP 0694070 A	31-01-1996
		FI 954601 A	23-02-1996
		JP 9503657 T	15-04-1997
		NO 953901 A	04-01-1996
		WO 9507994 A	23-03-1995
		US 6015686 A	18-01-2000
		US 5814482 A	29-09-1998
		US 5843723 A	01-12-1998
		US 5789245 A	04-08-1998
		US 6015694 A	18-01-2000
WO 9941397 A	19-08-1999	AU 2527499 A	30-08-1999